

GenCore version 5.1.1-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 17:32:29 ; Search time 2217 Seconds

(7487.765 Million cell updates/sec)

Title: US-09-515-363c-2

Perfect score: 5311

Sequence: 1 MSGCYSTDENFRYLISCFA.....LPTEPNLDSECCLEFSD 1025

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame_plus_p2n.model -DEV-xlh
-Q/cgn2_1/USPTO/US09515363/runat_16052003_100247_248/app_query.fasta.1.1223
-DB-EST-QFMT-fastlap -SUFIX-first -MINMATCH=0.1 -LDOPT=0 -L-PEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human4.cdi -LIST=45
-DOCALLIGN=200 -NORM SCORE-PCT -THR MAX=100 -THR MIN=0 -ALIGN=5 -MODE-LOCAL
-OUTFMT=oto -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER-US09515363.ecgn_1.1.1650_etunat_16052003_100247_24835 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Descr: Jon |
|------------|--------|-------|--------|-------|------------|
| 1 | 1597 | 30.1 | 1013 | 14 | BQ233683 |
| 2 | 1376 | 25.9 | 870 | 14 | BQ960157 |
| 3 | 1218 | 22.9 | 1115 | 13 | BM467983 |
| 4 | 1206.5 | 22.7 | 1035 | 12 | BF337464 |
| 5 | 1129 | 21.3 | 729 | 12 | BF983236 |
| 6 | 1122 | 21.1 | 1197 | 11 | AK018602 |
| 7 | 1098 | 20.7 | 827 | 12 | BG741146 |
| 8 | 1094.5 | 20.6 | 781 | 12 | BF686405 |
| 9 | 1081.5 | 20.4 | 1041 | 12 | BE882040 |
| 10 | 1066 | 20.1 | 1046 | 13 | BM476961 |
| 11 | 1059 | 19.9 | 755 | 14 | BQ772836 |
| 12 | 1055.5 | 19.9 | 1239 | 13 | BM467774 |
| 13 | 1016 | 19.1 | 672 | 14 | BQ316075 |
| 14 | 1016 | 19.1 | 672 | 14 | BQ316108 |
| 15 | 1016 | 19.1 | 672 | 14 | BQ316120 |
| 16 | 990.5 | 18.6 | 911 | 13 | BI454996 |
| 17 | 989.5 | 18.6 | 1611 | 13 | BM455238 |
| 18 | 952.5 | 17.9 | 932 | 13 | BC967530 |
| 19 | 911.5 | 17.2 | 568 | 14 | BQ311714 |
| 20 | 900 | 16.9 | 2398 | 11 | BC027369 |
| 21 | 896.5 | 16.9 | 790 | 12 | BF160649 |
| 22 | 892 | 16.8 | 1201 | 13 | BM551088 |
| 23 | 857.5 | 16.1 | 666 | 10 | BB148072 |
| 24 | 855 | 16.1 | 619 | 14 | BQ016388 |
| 25 | 841.5 | 15.8 | 787 | 13 | BI453288 |
| 26 | 836.5 | 15.8 | 596 | 12 | BF146945 |
| 27 | 830.5 | 15.6 | 715 | 10 | BB627907 |
| 28 | 827.5 | 15.6 | 761 | 13 | BI689931 |
| 29 | 813.5 | 15.3 | 585 | 9 | AA123960 |
| 30 | 796 | 15.0 | 546 | 12 | BE940626 |
| 31 | 779 | 14.7 | 493 | 12 | BC391418 |
| 32 | 774.5 | 14.6 | 481 | 10 | AM580584 |
| 33 | 765 | 14.4 | 1319 | 14 | BM806083 |
| 34 | 751 | 14.1 | 536 | 9 | AT171827 |
| 35 | 739.5 | 13.9 | 581 | 13 | BM426861 |
| 36 | 724 | 13.6 | 537 | 9 | A1806204 |
| 37 | 722 | 13.6 | 521 | 9 | A1922705 |
| 38 | 715 | 13.5 | 446 | 10 | BE090624 |
| 39 | 705 | 13.3 | 452 | 10 | BE090630 |
| 40 | 704 | 13.3 | 445 | 10 | BE090627 |
| 41 | 701 | 13.2 | 457 | 10 | BE090628 |
| 42 | 693 | 13.0 | 552 | 9 | AA134958 |
| 43 | 672.5 | 12.7 | 520 | 10 | BB284625 |
| 44 | 670 | 12.6 | 538 | 10 | AM189584 |
| 45 | 660 | 12.4 | 484 | 10 | AM152541 |

ALIGNMENTS

RESULT 1
BQ233683
LOCUS
DEFINITION BQ233683 1013 bp mRNA linear EST 02-MAY-2002
5', mRNA sequence.
ACCESSION BQ233683
VERSION BQ233683.1 GI:20415083
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1013)
NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abbs-remail.nih.gov

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM12875 row: 9 column: 07
 High quality sequence stop: 721.
 Location/Qualifiers

FEATURES

source
 1..1013
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5786526"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site-1: Not;
 Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb.
 BASE COUNT 383 a 161 c 229 g 240 t
 ORIGIN

Alignment Scores:

Pred. No.: 5,376-168 Length: 1013
 Score: 1597.00 Matches: 322
 Percent Similarity: 95.61% Conservative: 5
 Best Local Similarity: 94.15% Mismatches: 9
 Query Match: 30.07% Indels: 8
 DB: 14 Gaps: 3

US-09-515-363C-2 (1-1025) x BQ233683 (1-1013)

QY 529 LysAsnGlnIleGlnGluProCysLysLysPheAlaIleAlaSpaLathrArgGluasp 548
 |||||
 Db 2 AAAAACAATACAGAGCCATCGCAAGAACTTGCCATTCACATGCAACAGAGAGAT 61
 QY 549 ProPheLysGluLysLeuGluIleMetThrArgIleIleThrTyrCysIleMetSer 568
 |||||
 Db 62 CCATTTAAAGAGAACTTAGAAATATGACAAAGCATTAAACTTATGTCAGAAATGAT 121
 QY 569 PrometSeraphheGlyThrGlnProTyrGluGlnTrpAlaIleGlnMGIulysLys 588
 |||||
 Db 122 CCAATGTCAGATTTGGAACTCAACCTTAGAACAACTGCCCTTCAATAGCAAAAAA 181
 QY 589 AlaAlaLysLysLysAsnArgLysGluArgValCysAlaGluIleIleLysLeuTrpAsn 608
 |||||
 Db 182 GCTGCAAAAGCAAGAAATCGCAAGACGTTGTCATACCATTTGACAGAGTACAT 241
 QY 609 GluAlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaTyrThrIleLeuGluThr 628
 |||||
 Db 242 GAGGCCCTCAAAATTAATGACACAACTTCGAAATGATAGATACGATACCTGTTGAAT 301
 QY 629 PheTyrAsnGluLysLysAspLysPheAlaValIleGluAspAspLysAspGluGly 648
 |||||
 Db 302 TTCTAATAGTAAAGAAAGTAAAGATTGCGATCATAGCATATGATATGAGGCT 361
 QY 649 GluAspAspGluTyrCysAspGluAspGluAspGluAspAspLysLysProLeuLys 668
 |||||
 Db 362 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 QY 669 LeuAspGluThrAspArgPheLeuMetThrLeuPhePheGluAsnAsnLysMetLeuLys 688
 |||||
 Db 422 CTCGATGCAAAACATAGATTTCTCATGACCTTATTTTGAACAAATAAATGTTGAAA 481
 QY 689 ArgLeuAlaGluAsnProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIle 708
 |||||
 Db 482 AGCTGCTGCTGCAAAACCCAGATATGAAATGAAAGAACTGACCAATTAATAATACATA 541
 QY 709 MetGluGlnTyrThrArgThrGluGluGluSerAlaArgGlyIleIlePheThrLysThrArg 728

Db 542 ATGGAGCAATTAAGTACAGACTGAGAAATCAGCAGAGCAAGATTAATTTACAGAAACAGA 601
 QY 729 GlnSerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLysPheAlaIleValGly 748
 |||||
 Db 602 CAGAGTGCATATGCGCTTTCCAGTGCATTTACTGAAATGAAAAATTTGCTCAAGTACGA 661
 QY 749 ValLysAlaIleHisIleLeuIleGluValaGluIleHisSerSerGluPheLysProMet 768
 |||||
 Db 662 GTCAAAAGCCACCATCTGATGATGAGCTGACACAGCAGATGATGATGATGATGATGAT 721
 QY 769 AsnGluGlnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeuProIleAla 788
 |||||
 Db 722 AATGAACAAAGAAAGTCAATTAATTTCCAGCTGCAAAAAATTAATCTCTTAATCT 781
 QY 789 ThrThrValAlaGluGluLysLysLysLysLysLysLysLysLysLysLysLysLys 808
 |||||
 Db 782 ACCACAGTGCAGAAAGAAAGTGTGATTTAAAGATGATACATTTGATTCGTTATGCT 841
 QY 809 LeuValThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspSer 828
 |||||
 Db 842 CTCGCTACAAATGAATAGCCATGCTCAGGCCCTGCTGCGAGCCAGCTGATGAGAGC 900
 QY 829 ThrThrValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrVal 846
 |||||
 Db 901 ACCTACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
 QY 847 AsnAspPheArgGluLysMetMetTyrLysAlaIleHisCysValGlnAsnMetLysPro 866
 |||||
 Db 952 AATGATTTCCGAGAAGATGATGATTAAGCTATACATGCTGGCTTCAAAATGAAACG 1007
 QY 867 GluGlu 868
 |||||
 Db 1008 GAGGAG 1013

RESULT 2

LOCUS BQ960157 870 bp mRNA linear EST 21-AUG-2002
 DEFINITION AGENCOURT_8949486 NIH_MGC_71 Homo sapiens cdna clone IMAGE:6472264
 5', mRNA sequence.

ACCESSION BQ960157
 VERSION BQ960157.1 GI:22375635
 KEYWORDS EST.

SOURCE

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgs.nci.nih.gov/
 1 (bases 1 to 870)

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: c9abbs-remail.nih.gov

COMMENT Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: L1AM14005 row: 0 column: 17

High quality sequence stop: 738.
 Location/Qualifiers

1..870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="6472264"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site-1: Not;
 Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb.

Db 301 TACAGATAAACCATCACTAATACTTCTTGGCAAAACTGCAGTGTCTAGGCTGT 360
 QY 916 SerGluGluAspIleHisValIleGluLysMetHisHisValAsnMetThrProGluPhe 935
 Db 361 TCTGGGAGATATCCATGTAATGAGAAAATGCATCATCATATGATGATGAGAAATTC 420
 QY 936 LysGluLeuThrIleValArgGluAsnLysAlaLeuGlnLysLysGlySerAlaAspTyrGln 955
 Db 421 AAGGACCTTACATGTGTAAGAGAAACAAAACACACGCAAAAGTGTGTGACTATCAA 480
 QY 956 IleAsnGlyLysIleLeuLysCysGlyGlnAlaTyrPheThrMetMetValHisLys 975
 Db 481 ATAAATGGTAATCATCTGCAATATGCGCAGGCTGGGAGACATGATGATGCGACAAA 540
 QY 976 GlyLeuAspLeuProCysLeuLysIleArgAsnPheValValPheLysAsnAsnSer 995
 Db 541 GGCTTAACTTGGCTGTCTCCAAAATAAGCAATTTGTGTGCTTTCAAAAATAATCA 600
 QY 996 ThrLysGlnGlnTyrLysLysTyrValGluLeuProIleThrPheProAsnLeuAspTyr 1015
 Db 601 ACAAGAAACATACAAAAAGTGGTAAGATTACCTATTCATTCATTCACAACTTGACTAT 660
 QY 1016 SerGluCysLeuPheSerAspGluAsp 1025
 Db 661 TCAGAAATGCTGTTATTACTGATGAGAT 690

RESULT 4
 LOCUS BF337464 1035 bp mRNA line: EST 22-NOV-2000
 DEFINITION 602035195F1 NCI_CGAP_Brn64 Homo sapiens CDNA clone IMAGE:4181126
 ACCESSION BF337464
 VERSION BF337464.1 GI:11283715
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
 NIH-MGC http://mgi.mcl.nih.gov/
 1 (bases 1 to 1035)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rgsb@remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Place: LLNL9498 row: n column: 23
 High quality sequence start: 5
 High quality sequence stop: 695.
 Location/Qualifiers
 1..1035

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4181126"
 /clone.lib="NCI_CGAP_Brn64"
 /tissue.type="glio" (glioblastoma with EGFR amplification)
 /lab.host="DH10B (TI phage-resistant)"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.57 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP library."
 BASE COUNT 353 a 221 c 243 g 218 t

Alignment Scores:
 Pred. No.: 3..78e-124 Length: 1035
 Score: 1206.50 Matches: 264
 Percent Similarity: 84.04% Conservative: 15

Best Local Similarity: 79.52% Mismatches: 44
 Query Match: 22.72% Indels: 12
 DB: 12 Gaps: 5
 US-09-515-363C-2 (1-1025) x BF337464 (1-1035)
 QY 295 AsnValAlaIleArgAlaSerProGluLeuGlnLeuArgProTyrGln 314
 Db 6 AATGGCGGAGCAAGACATCCCGGAGCGAGAACTCCAGCTGAGCTTACTAAAGGAA 65
 QY 315 ValAlaGlnProAlaLeuGlnLysAsnIleIleLeuProThrLysGly 334
 Db 66 GTTGGCCAGCCAGCCCTTGAAGAGGAAATATCATCATCTGCTCTACAGGAAAGGA 125
 QY 335 LysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLysLysLysLys 354
 Db 126 AAACACAGATGGCTGTGTACATGTCAGAGATCATAGCAAGAGAAAAGCAATCT 185
 QY 355 GluProGlyLysValIleValLeuValAsnLysValLeuLeuValGluLeuPheArg 374
 Db 186 GAGCGTGGAAAGTATAGTCTTGTCAATAGGTAAGTGTGTAAGTGAACAGCTCTTGGC 245
 QY 375 LysGluPheGlnProPheLeuLysTyrTyrArgValIleGlyLeuSerLysThr 394
 Db 246 AAGGACTTCCAAACATTTTGAAGAAATGATGCTGTTATGATTAAGTGTATACAC 305
 QY 395 GlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleSerThrAla 414
 Db 306 CAACGTAAATATCATTTTCCAGAGCTGCAAGCTGCTGATATATTATCAGAAAGCT 365
 QY 415 GlnIleLeuGluAsnSerLeuLeuGlnGluGlnGluGlnGluGlnGluGlnGln 434
 Db 366 CAAATCTTGAAGAACTCCCTTTAACTTGAAGAAATGAGAGAGATGCTGTTCATTTG 425
 QY 435 SerAspPheSerLeuIleIleLeuAspGluCysHisIleThrAsnLysGlnAlaValTyr 454
 Db 426 TCAGACTTTCCTCATATATCATGATGATGATCATCACCAACAAAGAACATATAT 485
 QY 455 AsnAsnIleMetArgHisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLys 474
 Db 486 AATAACATCATGAGCATATATTGATGAGAAAGTGAAGAAACAAATAGACTCAACAAAGA 545
 QY 475 AsnLysProValIleProLeuProGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGln 494
 Db 546 AACAAACCAAGTATTCCTCTCCAGATGAGGAGCACTTCACTGCTGCTGCTGCTGCT 605
 QY 495 GlyAlaThrLysGlnAlaLysAlaGlnGlnHisIleLeuLysLeuCysAlaAsnLys 514
 Db 606 GGGCCACAGAACCAAGCCCAAACTGAAGACATTTTAAAGTATGTGCCAATTCAT 664
 QY 515 AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 534
 Db 665 GCATTTACTATTAATAAAGCTGTAAAGAAAACCTTGATTCACCTGAGAAACCAATAACA 724
 QY 535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGluLys 554
 Db 725 CCATCGCAAAAGTGGCGCT-GCGATGCGAC-CGAGAGATCCCTTAAGGAAAG- 779
 QY 555 LeuGlnIleMetThrArgIleGlnThrTyrCysGlnMetSerPro-MetSerAspIleG 574
 Db 780 ---TGTGAATATGGAGAGGTTCCACTTGTGTCAGT---ATGCCAGTCCGA---TTWGG 830
 QY 574 YThrGlnProTyrGlnGlnIleProAlaIleGlnMetGluLysLysAlaAlaLysLys 594
 Db 831 TAGCCACCTATTAA-----GGGCGCTCAACGGGGGAAAAAGGGCGCTAAAGAGAA 884
 QY 594 nArgLysGluArgValCysAlaGlnHisLeuArgLysTyrAsnGlnAlaLeuGlnHis 614
 Db 885 ACGGTGGGCGACCTTGGGCGCAACAGCGCACAAAGAGACAGAGAGAGAGAGAGAGAA 944
 QY 614 nAspThrIleArgMetIleAspAlaTyrThrHis 625
 Db 945 CGAACACAGCAAA-----GACAGAAAGCAACAGC 972

QY 982 LeuLysLeuArgAspPheValValValPheLysAsnAsnSerThrLysLysGlnTrpLys 1001
 DB 603 CTTAAATATAGGATTTGTAGTCAATTTCAAAATACCTACAGAGAAAGATGACAG 662
 QY 1002 TysTrpValGluLeuProIleThrPheProAsnLeuAspTyrSerGluGlyscysLeuPhe 1021
 DB 663 AAGTGGGTGGAAATGGCTTCAGATTTCTGATCTGACACTCAAAATACCTGCTGTAT 722
 QY 1022 SerAspGluAsp 1025
 DB 723 AGTGATGAAGAT 734
 RESULT 7
 LOCUS BG741146 827 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602631817p1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776869 5',
 mRNA sequence.
 ACCESSION BG741146
 VERSION BG741146.1 GI:14051799
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 827)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10630 row: b column: 06
 High quality sequence stop: 763.
 FEATURES
 source Location/Qualifiers
 1..827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4776869"
 /clone_lib="NCI_CGAP_Skn3"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: PCMV-SPOK3; Site_1: NCI;
 Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI-CGAP library."
 BASE COUNT 254 a 186 c 204 g 183 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,93e-112 Length: 827
 Score: 1098.00 Matches: 255
 Percent Similarity: 91.178 Conservative: 3
 Best Local Similarity: 90.118 Mismatches: 13
 Query Match: 20.674 Indels: 15
 DB: 12 Gaps: 0
 US-09-515-363c-2 (1-1025) x BG741146 (1-827)
 QY 87 GlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThrAspLeuProSerProSer 106
 DB 2 GAGGAGCCCTCTGGCCGC-CGCTACATGAACCCGAGCTCAGAGACTGCTCTCCATCG 60
 QY 107 PheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuAsnLeuGln-f:othrLeuVal 126
 DB 61 TTGGAGAACGCTCATGATGATATCTCCAACTCTCAACCTCCTCAGTCTCAGTCTGGT 120
 QY 126 LasPlyLeuValArgAspValLeuAspLysCysMetGluGluGlnLeuThrTrl 146

DB 121 GGACAGGCTTCTAGTTAGAGAGCTCTTGATTAAGTGCATGGAGGAGAACTGTGCAT 180
 QY 146 eGluAspArgAsnArgIleAlaAlaGluAsnAsnGlysnGlyserGlyValA-3G1 166
 DB 181 TGAACAGCAAAACCGGATTCGTGTCGAGAAACAAATGAAATGAAATGAGATGAA 240
 QY 166 uLeuLeuysArgIleValGlnLysGluAsnTrpPheSerAlaPheLeuAsnVal-104AT 186
 DB 241 GCTACTAAAGAGATTCGCAAGAAAGAAACGTTCTCTGCAATTCGATTCGATTCG 300
 QY 186 gGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrGlySerAsp--CysSer-1Ser 205
 DB 301 TCAACAGCAAAACATGAACTGTCTCAGAGATTACAGGCTCTGATTCGTCACAAAGC 360
 QY 206 AsnAlaGluIleGluAsnLeuSerGlnValAspGlyProGlnValGluGluLeu 225
 DB 361 AATGACAGATTCAGCAATTTATCACAGTTGATGCTCTCAAGTGAAGAGCACTTCT 420
 QY 226 SerThrThrValGluProAsnLeuGlu-LysGluValTrpGlyMetGluAsnSerSe 245
 DB 421 TCACCCACAGTTCCAGCAATCTGACGAGAGAGGTCTGGGCAATGAGAAATACCTATC 480
 QY 245 rGluSerSerPheAla-AspSerSerValValSerGluSerAspThrSerLeuAla-3G1 264
 DB 481 ACAATCATCTTTTGCACCATTTCTCTGATTCAGATTCAGATTCAGACAAAGTTTCCAT 540
 QY 264 uGlySerValSerCysLeuAspGluSerLeuGly-HisAsnSerAsnMetGlySerAsp 284
 DB 541 AGCAAGTGCACCTCTTATGATGAAAGTCTTGACCAATACGCAACATGGCTAGTAT 600
 QY 284 eGlyThrMetGlySerAspSerAspGlu-GluAsnValAlaAlaArgAlaSerProGlu 303
 DB 601 CAGGACCAATGGAGATGATTCAGATGAACGAAATGGGACGAGAGCAATCCCTGAG 660
 QY 304 ProGluLeuGlnLeuArgProTyrGlnMet--GluValAlaGlnProAlaLeuGluGly 323
 DB 661 CGAGAACTCCACCTCAGGCTTACCAAAATGGCAGAGTTGCCAGGCACG-TTGGCAAT 719
 QY 323 yAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrIle- 342
 DB 720 AGACTATCATTCATCTGCTC-CCTACAGGAGGAGGAGAAACAAACAGATCTGATTCAT 778
 QY 343 AlaLysAspHisLeuAspLysLysLysLysAlaSerGluProGly 357
 DB 779 GCCAGGATCACTTAGACAGAGAAACAAAGCATCTGAGCCTGGA 823
 RESULT 8
 LOCUS BF686405 781 bp mRNA linear EST 22-DEC-2000
 DEFINITION 602143786p1 NIH-MGC_46 Homo sapiens cDNA clone IMAGE:474805 5',
 mRNA sequence.
 ACCESSION BF686405
 VERSION BF686405.1 GI:11971813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 781)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM1170 row: 1 column: 22

High quality sequence stop: 731.

FEATURES
Location/Qualifiers

1..781

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4304805"

/issue_type="telomysarcoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 235 a 168 c 204 g 174 t

Alignment Scores:

Pred. No.: 8,83e-112

Score: 1094.50

Percent Similarity: 90.42%

Best Local Similarity: 90.04%

Query Match: 20,618

DB: 12

Gaps: 3

US-09-515-363c-2 (1-1025) x BE886405 (1-781)

74 GATTTTThrArgGluPheValGluAlaLeuArgThrGlySerProGluAlaAlaArg 94
8 GGTGGGAGCTGGGAAATTCGTGGAGGCCCTCGGAGAACCGCCCTCGGCCGCCGCG 67
94 TTTMetAsnProGluLeuThrAspLeuProSerProSerPheGluAsnAlaHisAspGlu 113
68 TACATGACCCCTGAGCTGACGAGCTGCGCTCCATCGTTGAGAACCTTATGATGAA 127
114 TTTLeuGlnLeuLeuAsnLeuLeuGlnProThrLeuValAspLysLeuValArgasp 133
128 TATCTCCAACTGCTGAACTCCCTTCAGCCCACTCTGTGTACAACTTCCTTACTAGAAC 187
134 ValLeuAspLysCysMetGluGluGluLeuLeuThrIleGluAspArgAsnAlaAla 153
188 GT-CTGATTAAGTCATGCGAGGAGACTGTTGACAAATTGAGACAGAAACCGATTGCT 246
154 AlaAlaGluAsnAsnGlyAsnGlyValArgGluLeuLeuLysArgIleValGln 173
247 GCTGCGAAACAAATGGAATGAAATGAGGTGTAAGAGCTTACTAAATGATTTGTCGAC 306
174 LysGluAsnTrpPheSerAlaPheLeuAsnValLeuArgIleThrGlyAsnGlnGlu 193
307 AAAGAAACCTGCTTCTCTGATTCGATTCTCTGCTGCTTCTGCTGCTTCTGCTGCTTCT 366
194 ValGlnGluLeuThrGlySerAspCysSerGluSerAspValGlnIleLeuAsnGln 213
367 GTCCAAAGACTTAACAGCGCTCTGATTCCTCGAAAGCAATTCAGCAATTCAGCAATTC 426
214 GlnValAspGlyProGlnValGluGlnLeuLeuSerThrThrValGlnProAsnGlu 233
427 CAAGTGATGGTCCCTCACTGAGAGCAACTTCTTCAACACAGACTTACGCAAAATCTG 486
234 GluLysGluValIleTrpGlyMetGluAsnAsnSerSerGluSerPheAlaAspSer 253
487 GAGAAAGAGCTGCGGCGCATGCAATTAATCATCATGAGAAATTCCTTTCAGATCTCT 546
254 ValValSerGluSerAspTrpSerLeuAlaGluGlySerValSerCysLysAspGluSer 273
547 GTAGTTTCAAGATCAGACCAAGTTGGGAGAGCAAGTTCAGCTTCAATCAAAAT 606
274 LeuGlyHisAsn-SerAspMetGlySerAspSerGlyThrMetGlySerAspSerAsp 293

607 CTTCGACATACAGACACACATGGCAGTATTCAGGACCACCATGGAGTATTTCATCA 666
293 uGluAsnValAlaAla---ArgAlaSerProGluProGluLeuGlnLeuArgProValGln 312
667 AGAGAAATGGGACGAGACATCCGAGACCAACCACTCA-AGCTTA-----CA 716
312 mMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleCysLysLeuProGln 332
717 AATGCAAGTGGCAGCAGCCTT-----GGAAGCAAGATATCATCATGCTCTCTCTCA 770
332 Y 332
771 G 771

RESULT 9
LOCUS BE882040
DEFINITION 601505326F2 NIH-MGC_71 Homo sapiens cDNA clone IMAGE:3065774 5',
mRNA sequence.

ACCESSION BE882040 1041 bp mRNA linear EST 20-OCT-2000
VERSION BE882040.1 GI:10330816
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1041)
NIH-MGC http://mgi.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: rstraus@remail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILM)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at:
http://image.llnl.gov

Plate: LAM9716 row: d column: 07
High quality sequence stop: 586.

FEATURES
Location/Qualifiers

1..1041

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:3906774"

/clone_lib="NIH-MGC_71"

/issue_type="telomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. primer: Oligo dT. Average insert size 2.1 kb."

BASE COUNT 385 a 258 c 223 g 175 t

ORIGIN

Alignment Scores:

Pred. No.: 4.1e-110

Score: 1081.50

Percent Similarity: 77.04%

Best Local Similarity: 72.96%

Query Match: 20,368

DB: 12

Gaps: 4

US-09-515-363c-2 (1-1025) x BE882040 (1-1041)

586 GluLysLysAlaAlaLysGlyAsnArgLysGluArgValCysAlaGluIleIleSerVal 605
2 GAAAAAAGAGCTGCAAAAGCAAAATGCAAGCAAGCTGTTGTGCGAGAAATTTCAGG 61
606 LysTyrAsnGluAlaLeuGlnIle-AsnAspThrIleArgMetIleAspAlaTyrIleHis 625
62 AAGTCAAGAGAGCGCTTCAAAATTAAGTACACATTCGATGATGATGCTGTAATTCATCA 121

QY 202 CysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAsp 216
|||||
Db 601 TCGTCAGAAAGCAATCGAGTATTGTAATTTACTAGAGAGAT 645
RESULT 11
B0772836/c 755 bp mRNA linear EST 26-JUL-2002
LOCUS B0772836
DEFINITION UI-H-FEO-bdn-c-04-0-UI_31 NCI_CGAP_FEO Homo sapiens cDNA clone
UI-H-FEO-bdn-c-04-0-UI_31, mRNA sequence.
ACCESSION B0772836
VERSION B0772836.1 GI:21981312
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 755)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Email: cgaaps-remail.nih.gov
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: James Martin
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@iowa.edu.
Seq primer: M13 FORWARD
POLY-A-Tes.
FEATURES
source Location/Qualifiers
1..755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-FEO-bdn-c-04-0-UI"
/clone_1lb="NCI_CGAP_FEO"
/tissue_type="Chondrosarcoma Cell Line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoR I; Site 2: Not I; NCI_CGAP_FEO is
a cDNA library containing the following tissue(s): a pool
of 3 chondrosarcoma cell lines (grade 2) The library was
constructed according to Bonaldo, Lennu and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into p773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCTACGAGC. The cell lines was provided by Dr James Martin
of University of Iowa.
TAG_LIB=UI-H-FEO
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_SEQ=CGCTACGAGC
BASE COUNT 207 a 141 c 115 g 290 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 8.06e-108 Length: 755
Score: 1059.00 Matches: 201
Percent Similarity: 98.53% Conservative: 0
Best Local Similarity: 98.53% Mismatches: 3
Query Match: 19.94% Indels: 1
DB: 14 Gaps: 0
US-09-515-363c-2 (1-1025) x B0772836 (1-755)
QY 822 ArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlyS-PolyValIle 841

Db 752 CGAGCCAGAGCTGATGAGAGGACCTAGCTCT-CTTGTCACTGTTTCAGCA-TAATC 694
|||||
QY 842 GluHisGluThrValAspPheArgGluLysMetMetTyrLysAlaIleHisTyrVal 861
|||||
Db 693 GACATGACAGACTTAATGATTTCCGAGAGAAATGATGATTAAGCATCATTTGGT 634
|||||
QY 862 GlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuMetGlnSerIle 881
|||||
Db 633 CAAATATGAAACCCAGAGAGATGCTCATAGATTGGAAATTAACGATGTAAAGATA 574
|||||
QY 882 MetGluLysLysMetLysThrLysArgSerLysLysHisTyrLysAsnAspProSer 901
|||||
Db 573 ATGCAAAAGAAATGAAAACCAAGAGAAATATGGCAAGCATTAACAAGATTAACCA 514
|||||
QY 902 LeuIleThrPheLeuCysLysAsnGlySerValLeuAlaCysSerGluLysIleHis 921
|||||
Db 513 CTAAATACCTTCCTTCGAAAAACGCGAGTGTGAGCTGTTCTGGGGAAGATATCAT 454
|||||
QY 922 ValIleGluLysMetHisValAsnMetThrProGluPheLysGluLeuTyrIleVal 941
|||||
Db 453 GTAATTGAGAAATGTCATCAGCTCATATGACCCAGAAATCAAGAACTTACATTTGTA 394
|||||
QY 942 ArgGluAsnLysAlaLeuGluLysLysCysAlaAspTyrGlnIleAsnGluIleIle 961
|||||
Db 393 AGAGAAACAAAGCAGCTCAAAAGAGTGTCCGACTTCATTAATTAATGATGAATATATC 334
|||||
QY 962 CysLysCysGlyGlnAlaIleTyrGlyThrMetMetValHisLysGlyLeuAspPheProLys 981
|||||
Db 333 TGCAAATGTGGCGACAGCTTGGGGAACATGATGTGGCAAAAGCTTAGATTTCCTTGT 274
|||||
QY 982 LeuLysIleArgAsnPheValValIlePheLysAsnAsnSerThrLysLysGlnTyrLys 1001
|||||
Db 273 CTCAAATATAGAAATTTGTAGTGTTCACAAATATTCACAAAGAAACAAACAA 214
|||||
QY 1002 LysTPPValIleGluLeuProIleThrPheProAsnLeuAspTyrSerGluCysCysLysPhe 1021
|||||
Db 213 AAGTGGGAGAAATTAACCTATCATATTTCCCAATCTTGACTTATTAAGAAATGCTGTTATTT 154
|||||
QY 1022 SerAspGluAsp 1025
|||||
Db 153 AGTGTAGAGAT 142
|||||
RESULT 12
B0467774 1239 bp mRNA linear EST 05-FEB-2002
LOCUS B0467774
DEFINITION AGNCOURT 6437783 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532660
5', mRNA sequence.
ACCESSION B0467774
VERSION B0467774.1 GI:18516816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1239)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNI)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNI at:
<http://image.lni.gov>
Plate: LAM12216 row: e column: 13
High quality sequence stop: 505.
FEATURES
source Location/Qualifiers
1..1239
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:5532660"
/clone_lib="NH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-Sport5; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      393 a      285 c      243 g      318 t
ORIGIN
Alignment Scores:
Pred. No.:      4,53e-107      Length:      1239
Score:          1055.50      Matches:      229
Percent Similarity: 86.91%      Conservative: 10
Best Local Similarity: 83.27%      Mismatches: 23
Query Match:     19.87%      Indels:      13
DB:              Gaps:      4
US-09-515-363c-2 (1-1025) x BM467774 (1-1239)
QY      190 AsnAsnGluLeuValGlnGluLeuThrGlySerAspCysSerGlnSerAlaGluIle 209
      |||||
      4 AACAAATGAAGCTGCTCCAGAGTTAAACAGGCTGATGCTCAGAAAGCAATGCAGAGATT 63
QY      210 GluAsnLeuSerGlnValAspGlyProGlnValGluGlnLeuLeuSerThrIleVal 229
      |||||
      64 GAAATATTATCACAAATGTAAGTCTCTCAAGTGAAGAGCAACTTCTTTTAAACATAGATT 123
QY      230 GluProAsnLeuGluLysGluValTyrGlyMetGluAsnAsnSerSerGlnSerSerPhe 249
      |||||
      124 CACCAAAATCTGAGAGAGAGAGTCTGGGCGATGAGAAATACATCATCATCATCTTTT 183
QY      250 AlaAspSerSerValValSerGlnSerAspThrSerLeuAlaGluLysValSerGln 269
      |||||
      184 GCAGATTCTCTCTAGTTCATTCAGAAATGAGACAAATTTGGCAGAGAGAAATGCAGCTTC 243
QY      270 LeuAsnGluSerLeuGlnLysAsnSerAsnMetGlySerAspSerGlyTyrMetGlySer 289
      |||||
      244 TTAGATGAAGAGTCTTGACATTAACACCAACATGGCAGATGATTCAGGCAATATGAGT 303
QY      290 AspSerAspGluGluAsnValAlaAlaArgAlaSerProGlnProGlnGlnLeuVal 309
      |||||
      304 GATTCAATGTAAGAGATGAGAGAGAGAGATCCCGGAGGAGCAATCTCCAGCTCAG 363
QY      310 ProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleCysLeu 329
      |||||
      364 CCTTACCAATGAGAGATGGCCAGCCAGCTTGGAAGGAGAGAAATCATCATCTGCTC 423
QY      330 ProThrGlySerGlyLysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLys 349
      |||||
      424 CCAACAGGAGGTGGAAGAACAGAGAGGCTGTTACATTCACAGATCATCATAGCAAG 483
QY      350 LysLysLysAlaSerGlnProGlyLysValIleValLeuValAsnLysValLeuLeuVal 369
      |||||
      484 AAGAAAAAGCATCTGAGCCTGGAAGATTATAGCTCTTTCATTAAGTACAGTCAAGTT 543
QY      370 GluGlnLeuPheArgLysGluPheGlnPro-PheLeuLysLysTyrTyrArgValIleGln 389
      |||||
      544 GAACAGCTCTCTCCGAGAGAGATCCCAACCTTTTGAAGAAAGGACATCTGTTATGCG 603
QY      389 LysSer-GlyAspThrGlnLeuLysIleSerPheProGlnValValLys-SerCysAsp 408
      |||||
      604 ATTAAAGGGGTATACCAACCAATATCATTTCCACAGATGTTCAAATTCCTGTCAT 663
QY      409 IleIle-----IleSerThrAlaGlnIleLeuGluAsnSer-LeuLeuAsnLeu---GI 425
      |||
      664 ATTATATATCATCAAGCTCAAAATTCCTTTGAAAAAAGCCGCCCTTAAAGCTTGAGAA 723
QY      425 uaaanglygluasp-----AlaGlyValGlnLeuSerAspPheSerLeuIleIleIleAs 443
      |||||
      724 AATGGGAGAAAAACGCTTGGCGGCTTCCAACTCCCAAGAACTTTTCCCTTCATT----- 778

```

```

QY      443 pGluCysHisHisThrAsnLysGluAlaValTyrAsn 455
      |||||
      DB      779 -----TATCATTTGATGAAGAAATGTTATCATCCACAC 808
      |||||
      RESULT 13
      LOCUS      BQ316075/C
      DEFINITION  BQ316075 672 bp mRNA linear EST 17-MAY-2002
      ACCESSION  CM3-CT0275-191099-024-909 CT0275 Homo sapiens cDNA. mRNA sequence.
      VERSION    BQ316075
      KEYWORDS   BQ316075.1 GI:20921844
      SOURCE     EST.
      ORGANISM   human.
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae;
      Homininae; Homo; Homo sapiens; Homo sapiens; Homo sapiens; Homo sapiens;
      1 (bases 1 to 672)
      Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Bittencourt, M.R.,
      Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
      Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
      Brunstein, A., de Oliveira, P.S., Bucher, P., Jomeneel, V., O'Hare,
      M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S. and
      Simpson, A.J.G.
      Shotgun sequencing of the human transcriptome with cDNA expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-3704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=CM3&ct=CM3-CT0275-
      191099-024-909&ct=1999-10-19&ct=1)
      Seq primer: puc 18 forward
      High quality sequence start: 50
      High quality sequence stop: 671.
      Location/Qualifiers
          1..672
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="CT0275"
             /dev_stage="Adult"
             /note="Organ: colon; Vector: puc18; Site: 1; Site: 2:
             SmaI; A mini-library was made by cloning products derived
             from ORESTES PCR (U.S. Letters Patent application No. 196
             716 - Ludwig Institute for Cancer Research) profiles
             into the puc 18 vector. Reverse transcription of tissue
             mRNA and cDNA amplification were performed under low
             stringency conditions."
BASE COUNT      165 a      158 c      120 g      229 t
ORIGIN
Alignment Scores:
Pred. No.:      4.46e-103      Length:      672
Score:          1016.00      Matches:      202
Percent Similarity: 94.62%      Conservative: 9
Best Local Similarity: 90.58%      Mismatches: 10
Query Match:     19.13%      Indels:      2
DB:              Gaps:      2
US-09-515-363c-2 (1-1025) x BQ316075 (1-672)
QY      652 GluTyrCysAspGlyAspGluAspGluAspGluAspGluAspGluAspGlu 671
      |||||
      DB      671 GAGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 612
      |||||
      QY      672 ThrAspArgPheLeuMetThrLeuPheGlnGluAsnAsnLysMetLeuLysArgValAla 691
      |||||

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:48:28 ; Search time 48 Secs/nds
(without alignments)
4399.964 Million cell updates/sec

Title: US-09-515-363c-2

Perfect score: 5311
Sequence: 1 MSNGYSTDENFRYLISCFRA.....LPITFVNDYSEICLFSDSD 1025

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rotent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 5311 | 100.0 | 1025 | 4 Q9H3G6 | Q9H3G6 homo sapien |
| 2 | 5285 | 99.5 | 1025 | 4 Q9BYX4 | Q9BYX4 homo sapien |
| 3 | 4199 | 79.1 | 1025 | 11 Q9R5F7 | Q9R5F7 mus musculu |
| 4 | 2456 | 46.2 | 468 | 4 Q9EMX8 | Q9EMX8 homo sapien |
| 5 | 2271.5 | 42.8 | 514 | 11 Q99KS4 | Q99KS4 mus musculu |
| 6 | 2059.5 | 38.8 | 467 | 11 Q9R144 | Q9R144 mus musculu |
| 7 | 1356 | 25.5 | 678 | 11 Q99J87 | Q99J87 mus musculu |
| 8 | 1352 | 25.3 | 678 | 4 Q96C10 | Q96C10 homo sapien |
| 9 | 1344 | 25.3 | 678 | 4 Q9HAM6 | Q9HAM6 homo sapien |
| 10 | 1086.5 | 20.5 | 925 | 4 Q95786 | Q95786 mus musculu |
| 11 | 1018 | 19.2 | 940 | 6 Q9GLV6 | Q9GLV6 mus scrofa |
| 12 | 960 | 18.1 | 210 | 11 Q9D225 | Q9D225 mus musculu |
| 13 | 687 | 12.9 | 1037 | 5 Q44165 | Q44165 caenorhabdi |
| 14 | 678 | 12.8 | 143 | 11 Q9VED9 | Q9VED9 mus musculu |
| 15 | 654 | 12.3 | 811 | 5 Q17545 | Q17545 caenorhabdi |
| 16 | 516 | 9.7 | 398 | 4 Q9MT04 | Q9MT04 homo sapien |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 17 | 499 | 9.4 | 1134 | 5 Q93413 | Q93413 caenorhabdi |
| 18 | 453 | 8.5 | 752 | 17 Q9V1Z5 | Q9V1Z5 pyrococcus |
| 19 | 441.5 | 8.3 | 650 | 17 Q95524 | Q95524 pyrococcus |
| 20 | 417.5 | 7.9 | 741 | 17 Q28614 | Q28614 archaeoglob |
| 21 | 408 | 7.7 | 1909 | 10 Q9SP32 | Q9SP32 arabidopsis |
| 22 | 406 | 7.6 | 1909 | 10 Q9EPY6 | Q9EPY6 arabidopsis |
| 23 | 402.5 | 7.6 | 2024 | 10 Q9MAN0 | Q9MAN0 arabidopsis |
| 24 | 395.5 | 7.4 | 764 | 17 Q8TZH8 | Q8TZH8 pyrococcus |
| 25 | 389 | 7.3 | 821 | 17 Q8TUH1 | Q8TUH1 methanosarc |
| 26 | 354 | 6.7 | 1539 | 11 Q8K418 | Q8K418 mus musculu |
| 27 | 350 | 6.6 | 513 | 17 Q978A0 | Q978A0 thermoplasma |
| 28 | 343 | 6.5 | 182 | 11 Q9D1X4 | Q9D1X4 mus musculu |
| 29 | 343 | 6.5 | 2042 | 10 Q9M9P8 | Q9M9P8 arabidopsis |
| 30 | 342 | 6.4 | 738 | 17 Q27466 | Q27466 methanobact |
| 31 | 332.5 | 6.3 | 1383 | 5 Q9VDA0 | Q9VDA0 drosophila |
| 32 | 332 | 6.3 | 508 | 17 Q9H146 | Q9H146 thermoplasma |
| 33 | 332 | 6.3 | 1924 | 4 Q9UPY3 | Q9UPY3 homo sapien |
| 34 | 313.5 | 5.9 | 784 | 17 Q9HMM5 | Q9HMM5 haemobacteri |
| 35 | 309.5 | 5.8 | 1883 | 5 Q952G7 | Q952G7 dictyostell |
| 36 | 302.5 | 5.7 | 741 | 17 Q8TUS6 | Q8TUS6 methanopyru |
| 37 | 302.5 | 5.7 | 1789 | 5 Q8T2B3 | Q8T2B3 dictyostell |
| 38 | 295.5 | 5.6 | 783 | 3 Q9HE09 | Q9HE09 schizosacch |
| 39 | 295 | 5.6 | 1458 | 10 Q8W367 | Q8W367 oryza sativ |
| 40 | 293.5 | 5.5 | 1722 | 5 Q95Y33 | Q95Y33 drosophila |
| 41 | 286 | 5.4 | 503 | 1 Q74070 | Q74070 cenarchaeum |
| 42 | 284 | 5.3 | 502 | 1 Q74070 | Q74070 cenarchaeum |
| 43 | 278.5 | 5.2 | 839 | 3 Q9UT23 | Q9UT23 schizosacch |
| 44 | 278.5 | 5.2 | 1540 | 3 Q8X079 | Q8X079 neurospora |
| 45 | 258.5 | 4.9 | 1587 | 10 Q91OE5 | Q91OE5 arabidopsis |

ALIGNMENTS

RESULT 1
Q9H3G6 PRELIMINARY; PRT: 1025 AA.
ID Q9H3G6: Q9H3G6:
AC Q9H3G6:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Melanoma differentiation associated protein-5.
GN MDA5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Kang D.-C., Fisher P.B.;
RT *Melanoma differentiation associated gene-5 (mda-5), an intericton
RT inducible gene of limited homology to RNA helicase.*;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF095844; AAC34368.1;
DR InterPro: IPR001315; CARD.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD_1.
DR Pfam: PF00271; helicase_C_1.
DR SMART: SM00487; DEHC_1.
DR SMART: SM00490; HELIC_1.
DR PROSITE: PS0209; CARD_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1025 AA; 116687 MW; 48BE75491D863741 CRC64;
Query Match: 100.0%; Score 5311; DB 4; Length 1025.
Best Local Similarity 100.0%; Pred. NO.1.8e-283;
Matches 1025; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSNGYSTDENFRYLISCFRAKMYIQVEPYDLTFPLPAVKKEQIQRTVATSGMMQAVE 60
|||||
Db 1 MSNGYSTDENFRYLISCFRAKMYIQVEPYDLTFPLPAVKKEQIQRTVATSGMMQAVE 60

```

OY 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYLOLNL 120
DB 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYLOLNL 120
OY 121 LQPTLVADKLLVROVLDKCMEEELLTIEDNRNRIAAENNGNSVPELTK-VOKERENFA 180
DB 121 LQPTLVADKLLVROVLDKCMEEELLTIEDNRNRIAAENNGNSVPELTK-VOKERENFA 180
OY 181 FLNVLROGTGNELVOELTSGDCSESNAEIENLSQVDPQVEEJLSTTVJ-NLEKIVWGM 240
DB 181 FLNVLROGTGNELVOELTSGDCSESNAEIENLSQVDPQVEEJLSTTVJ-NLEKIVWGM 240
OY 241 ENNSSESPADSSVYSESDTSLAEGSVSCDESLGHNSNGMSISGTMGS-SDDEENVAARA 300
DB 241 ENNSSESPADSSVYSESDTSLAEGSVSCDESLGHNSNGMSISGTMGS-SDDEENVAARA 300
OY 301 SPEPELOLRPYOMEVAQPALEGKNIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
DB 301 SPEPELOLRPYOMEVAQPALEGKNIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
OY 361 VLVNKKVLLVEOLFRKEFOPFLKKWRYVIGLSGTOLKISPEVYKSCDII-STAOJLENS 420
DB 361 VLVNKKVLLVEOLFRKEFOPFLKKWRYVIGLSGTOLKISPEVYKSCDII-STAOJLENS 420
OY 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYLMOKLKNN-IKKNRPVIR 480
DB 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYLMOKLKNN-IKKNRPVIR 480
OY 481 LPOLIGLTASPGVGATKOAKAEHILKLCANLDAFTIKYKKNLDOLKX-IOEPCPKFA 540
DB 481 LPOLIGLTASPGVGATKOAKAEHILKLCANLDAFTIKYKKNLDOLKX-IOEPCPKFA 540
OY 541 IADATREDPFKEKLEIMTRITOTYCOMSPMSDFGTOPYEOMAIOMEKKAAGKREKSVG 600
DB 541 IADATREDPFKEKLEIMTRITOTYCOMSPMSDFGTOPYEOMAIOMEKKAAGKREKSVG 600
OY 601 AEHLRKYNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIEDSDSEG-DEYCGDDHDE 660
DB 601 AEHLRKYNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIEDSDSEG-DEYCGDDHDE 660
OY 661 DDLKPKLIDETDRFLMTLEFFENNMKMLKLAENPEYENKLTLRNTIMH-QYTRTESAR 720
DB 661 DDLKPKLIDETDRFLMTLEFFENNMKMLKLAENPEYENKLTLRNTIMH-QYTRTESAR 720
OY 721 GIIFFTROSAYALSMITENEFKFAVYKVAHNLIGAGH-SEFKEMTONH-QKEVISKEPT 780
DB 721 GIIFFTROSAYALSMITENEFKFAVYKVAHNLIGAGH-SEFKEMTONH-QKEVISKEPT 780
OY 781 GKINLIATTVAEGLDIEKCNIVIRYGLVTNETIAYVOAKGRARADESTYVIVAHSGSV 840
DB 781 GKINLIATTVAEGLDIEKCNIVIRYGLVTNETIAYVOAKGRARADESTYVIVAHSGSV 840
OY 841 IEHETVNDERREKKMYKAIHCYONMKPREYAHKILLOMOSIMEKKKTKENIAKHKNNP 900
DB 841 IEHETVNDERREKKMYKAIHCYONMKPREYAHKILLOMOSIMEKKKTKENIAKHKNNP 900
OY 901 SLTTFELCKMCSYIACGEDIHYIEKHHYNNTPREKELYIVBENKALOKK-ADYOJNGEI 960
DB 901 SLTTFELCKMCSYIACGEDIHYIEKHHYNNTPREKELYIVBENKALOKK-ADYOJNGEI 960
OY 961 ICKGQAMGTIMVHKGDLDPCLKIRNFVYVFNKNNSTKKQKKMVELPITFIMLDYSECC 1020
DB 961 ICKGQAMGTIMVHKGDLDPCLKIRNFVYVFNKNNSTKKQKKMVELPITFIMLDYSECC 1020
OY 1021 FSDED 1025
DB 1021 FSDED 1025

```

```

AC 09BYX4:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE RNA helicase-DEAD box protein RH116.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Coccone C., Kolesnichenko V., Billaut-Mulot O., Truong M.-J.,
RA Capton A., Bahr G.M.;
RT Identification of a new RNA helicase (RH116) regulated by the
RT Immunomodulator Murabutide.
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY017378; AAG54076.1;
DR InterPro: IPR001315; CARD.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD_1.
DR Pfam: PF00271; Helicase_C_1.
DR SMART: SM00487; DEXDC_1.
DR SMART: SM00490; HELICG_1.
DR PROSITE: PS50209; CARD_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1025 AA; 116671 MW; 6B4F3D14E88A7D6D CRC64;

Query Match          99.5%; Score 5285; DB 4; Length 1025;
Best Local Similarity 99.5%; Pred. No. 4,9e-282;
Matches 1020; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSGNSTDENFRYLIISCFRARKMYIOVEPVLDVLTFLPAEYKEQIQRIATVSCNMCAVE 60
DB 1 MSGNSTDENFRYLIISCFRARKMYIOVEPVLDVLTFLPAEYKEQIQRIATVSCNMCAVE 60
OY 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYLOLNL 120
DB 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYLOLNL 120
OY 121 LQPTLVADKLLVROVLDKCMEEELLTIEDNRNRIAAENNGNSVPELTK-VOKERENFA 180
DB 121 LQPTLVADKLLVROVLDKCMEEELLTIEDNRNRIAAENNGNSVPELTK-VOKERENFA 180
OY 181 FLNVLROGTGNELVOELTSGDCSESNAEIENLSQVDPQVEEJLSTTVJ-NLEKIVWGM 240
DB 181 FLNVLROGTGNELVOELTSGDCSESNAEIENLSQVDPQVEEJLSTTVJ-NLEKIVWGM 240
OY 241 ENNSSESPADSSVYSESDTSLAEGSVSCDESLGHNSNGMSISGTMGS-SDDEENVAARA 300
DB 241 ENNSSESPADSSVYSESDTSLAEGSVSCDESLGHNSNGMSISGTMGS-SDDEENVAARA 300
OY 301 SPEPELOLRPYOMEVAQPALEGKNIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
DB 301 SPEPELOLRPYOMEVAQPALEGKNIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
OY 361 VLVNKKVLLVEOLFRKEFOPFLKKWRYVIGLSGTOLKISPEVYKSCDII-STAOJLENS 420
DB 361 VLVNKKVLLVEOLFRKEFOPFLKKWRYVIGLSGTOLKISPEVYKSCDII-STAOJLENS 420
OY 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYLMOKLKNN-IKKNRPVIR 480
DB 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYLMOKLKNN-IKKNRPVIR 480
OY 481 LPOLIGLTASPGVGATKOAKAEHILKLCANLDAFTIKYKKNLDOLKX-IOEPCPKFA 540
DB 481 LPOLIGLTASPGVGATKOAKAEHILKLCANLDAFTIKYKKNLDOLKX-IOEPCPKFA 540
OY 541 IADATREDPFKEKLEIMTRITOTYCOMSPMSDFGTOPYEOMAIOMEKKAAGKREKSVG 600
DB 541 IADATREDPFKEKLEIMTRITOTYCOMSPMSDFGTOPYEOMAIOMEKKAAGKREKSVG 600
OY 601 AEHLRKYNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIEDSDSEG-DEYCGDDHDE 660

```

```

Db 601 AEHLRYKNAQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGSEIEXYCDDEJE 660
QY 661 DDLAKKPLKLDDETRFLMTLFFENNKKMLKRLAENPEYENKSLTKLRNTIMQYTRTESAR 720
Db 661 DDLAKKPLKLDDETRFLMTLFFENNKKMLKRLAENPEYENKSLTKLRNTIMQYTRTESAR 720
QY 721 GILFTKROSAVALSOMITENEFKAEYGVAKHHLIGAGHSSEKPMPTONQKQVISEKRT 780
Db 721 GILFTKROSAVALSOMITENEFKAEYGVAKHHLIGAGHSSEKPMPTONQKQVISEKRT 780
QY 781 GKINLLIATTAEEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSSGV 840
Db 781 GKINLLIATTAEEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSSGV 840
QY 841 IEHEYNDFREKKMYKAIHCVONMKPEYAKHLLILOMOSIMEKKMKTKKINAKHYKNP 900
Db 841 IEHEYNDFREKKMYKAIHCVONMKPEYAKHLLILOMOSIMEKKMKTKKINAKHYKNP 900
QY 901 SLITFLCKNCVSLACSGEDIHVEIEKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 960
Db 901 SLITFLCKNCVSLACSGEDIHVEIEKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 960
QY 961 ICKGQAMGTMMVHKGDLPCIKIRNFVVFKNNSKTKQYKAWELPITFENLDSPECL 1020
Db 961 ICKGQAMGTMMVHKGDLPCIKIRNFVVFKNNSKTKQYKAWELPITFENLDSPECL 1020
QY 1021 FSDSD 1025
Db 1021 FSDSD 1025

RESULT 3
08R5F7 PRELIMINARY; PRT; 1025 AA.
AC 08R5F7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE HELICARD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kovacsovich M., Hofmann K., Tschopp J.;
RT "HELICARD, a novel CARD-containing helicase.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY075132; AAL77205.1; -.
SQ SEQUENCE 1025 AA; 115971 MW; 708FCAC690CFF6D8 CMC61;

Query Match 79.1%; Score 4199; DB 11; Length 1025;
Best Local Similarity 79.7%; Pred. No. 2.2e-22;
Matches 818; Conservative 75; Mismatches 131; Indels 2; Gaps 2;

```

```

Db 241 DDLPEASTDSSVTSIESDTSLAEGSVSCFDESLGNSNMKGDSGTMSDSESTYQIKR 300
QY 300 ASPEPELQRPYOMEVAQALGSKNIIICLPQSGCTRAVYIADHLDKKAASEYKAV 359
Db 301 VSPEPELQRPYOMEVAQALGSKNIIICLPQSGCTRAVYIADHLDKKAASEYKAV 360
QY 360 IYLVNVLVEOLFREPEPPLKMYRVIGLSGDPOLKISPEPVKSCDIIISTQOILEN 419
Db 361 IYLVNVLVEOLFREPEPPLKMYRVIGLSGDPOLKISPEPVKSCDIIISTQOILEN 420
QY 420 SLINLENGEDAGVQSLDSFLIIIDECHTNKSAVYNNIMRHLYMOKLKNRKLKKNKPVY 479
Db 421 SLINLENGEDAGVQSLDSFLIIIDECHTNKSAVYNNIMRHLYMOKLKNRKLKKNKPVY 480
QY 480 PLPQILGLTASPGVGCATQAKAEHILKCANLDAFTIKYKENLQDKNOIQEPCKE 539
Db 481 PLPQILGLTASPGVGCATQAKAEHILKCANLDAFTIKYKENLQDKNOIQEPCKE 540
QY 540 ATADATREDPPEKLEIMTRIOYQOMSPMSDQTOPYQOMAIQOMEKKAAGKRRKRY 599
Db 541 VIADDTREMPFEKLEIMASIOYQOMSPMSDQTOPYQOMAIQOMEKKAAGKRRKRY 600
QY 600 CAEHLRYKNAQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGDEYCDGDEY 659
Db 601 CAEHLRYKNAQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGDEYCDGDEY 659
QY 660 EDDLKAPKLDDETRFLMTLFFENNKKMLKRLAENPEYENKSLTKLRNTIMQYTRTESA 719
Db 660 EDDLKAPKLDDETRFLMTLFFENNKKMLKRLAENPEYENKSLTKLRNTIMQYTRTESA 719
QY 720 RGILFTKROSAVALSOMITENEFKAEYGVAKHHLIGAGHSSEKPMPTONQKQVISEKRT 779
Db 720 RGILFTKROSAVALSOMITENEFKAEYGVAKHHLIGAGHSSEKPMPTONQKQVISEKRT 779
QY 780 TGRINLLIATTAEEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSSGV 839
Db 780 TGRINLLIATTAEEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSSGV 839
QY 840 VIEHEYNDFREKKMYKAIHCVONMKPEYAKHLLILOMOSIMEKKMKTKKINAKHYKNP 899
Db 840 VIEHEYNDFREKKMYKAIHCVONMKPEYAKHLLILOMOSIMEKKMKTKKINAKHYKNP 899
QY 900 PSILITFLCKNCVSLACSGEDIHVEIEKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 959
Db 900 PSILITFLCKNCVSLACSGEDIHVEIEKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 959
QY 960 IICKGQAMGTMMVHKGDLPCIKIRNFVVFKNNSKTKQYKAWELPITFENLDSPECL 1019
Db 960 IICKGQAMGTMMVHKGDLPCIKIRNFVVFKNNSKTKQYKAWELPITFENLDSPECL 1019
QY 1020 LNSDED 1025
Db 1020 LNSDED 1025

RESULT 4
096MX8 PRELIMINARY; PRT; 468 AA.
AC 096MX8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CDNA FLJ31731 f15; clone NT2RT2006855, weakly similar to Homo sapiens
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., Uno Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,

```

| | | | | |
|-----------------------|-----------------|--------------------|----------|------------|
| Query Match | 38.8% | Score 2059.5 | DB 11 | Length 467 |
| Best Local Similarity | 84.6% | Pred. No. 2.4e-105 | | |
| Matches 396 | Conservative 28 | Mismatches 43 | Indels 1 | Gaps 1 |

```

Db 1 MASIQYQCKSPKSDFGTGYEOMAIOMKKAQDGNRRCORCAEHLKRYNEALQINDI 60
OY 618 RMIDATHTLETFYNEKKDKFAVIEDSDGDEYCDGJEDDOLKRLKLEIDRELM 677
Db 61 RMIDASHLETFYTDKEKFAVL-NDSDSDDEASSCQKQGVKSKIKLEIDRELM 119
OY 678 TLEFENKMLKRLAENPEYENKLTKLRTIMEOYTRIEASAGLIEFKIKOSAYALSNQ 737
Db 120 NLEFQKKMLKRLAENPKYENKLTKLRTIMEOYTRIEASAGLIEFKIKOSAYALSNQ 179
OY 738 ITENEKFAEYGVKAHHLIGAGHSEKPYMTONEQEVISKFRIGKTNLLIATVAESID 797
Db 180 IMENAFAYGVKAHHLIGAGHSEKPYMTONEQEVISKFRIGKTNLLIATVAESID 239
OY 798 IKCNIVIRGLVTNEIAYOARGARADESTVILVAHSISVIEHTVNDPREKKMYKA 857
Db 240 IKCNIVIRGLVTNEIAYOARGARADESTVILVAHSISVIEHTVNDPREKKMYKA 299
OY 858 INCVQMKPEEYVAKHILLOMOSIMEKKTKENIAKHYKNPSLITELKNCVLAESG 917
Db 300 INCVQMKPEEYVAKHILLOMOSIMEKKTKENIAKHYKNPSLITELKNCVLAESG 359
OY 918 EDIHYEKKHVMNTPFEKELTYVRENKALQKKCADYQNGETICKGQAGMTVMYKSL 977
Db 360 ENIHVIEKKHVMNTPFEKELTYVRENKALQKKCADYQNGETICKGQAGMTVMYKSL 419
OY 978 DLPCILIRNFVYVFNKNSPKKQKKNVLPITFPNLDYSACCLFSESD 1025
Db 420 DLPCILIRNFVYVFNKNSPKKQKKNVLPITFPNLDYSACCLFSESD 167

RESULT 7
OY 99J87 PRELIMINARY: PRT: 678 AA.
AC 099J87;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 76.7 kDa protein.
GN DILGPZE OR DILGPZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-1161808;
RA Miyoshi K., Cui Y., Riedinger G., Lehoczeky J., von L., Oka T.,
RA Dewar K., Hennighausen L.;
RT Structure of the Mouse Stat 3/5 Locus: Evolution from Protophila to
RT Zebrafish to Mouse.
RL Genomics 71:150-155(2001).
DR EMBL; AF317000; AAK15475.1; -
DR EMBL; AF316999; AAK15474.1; -
DR MGD; MG1:1931560; DILGPZE.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR ATP-binding; Helicase; Hypothetical protein.
SQ SEQUENCE 678 AA: 76726 MW: DC42B75A3AD37A68 JRC64:

Query Match 25.5%; Score 1356; DB 11; Length 678;
Best Local Similarity 41.7%; Pred. No. 1.7e-66;
Matches 299; Conservative 129; Mismatches 237; Indels 52; Gaps 12;
OY 306 IOIRPYOMEVAPALGKNNIICLPGSGKTRAVVIAKHILKKKKASIKGVITLVNK 365
Db 1 MEIRPYOMEVAPALGKNNIICLPGSGKTRAVVIAKHILKKKKASIKGVITLVNK 56
OY 366 VLLVEOLFKEFOPFLKKWYRIGLSGDTOLKISPEVVKSCIIISTALLENSLNIE 425

```

```

Db 57 VHLVSO-HAEFRRLDKHMTVTTLTSGMGSAGGLMARSHDLICTAELQIVINSE 115
OY 426 NQEDAGVQLSDSLIITDECHTNKEAYNNMRTYIMOKLKNRLKENK.VIII.POL 485
Db 116 --EDERVELREPSLIVNVECHTHKDTYNTLLSYREOKLK---KAE-----P 164
OY 486 GLTASPGVAGATKQAKAEHILKLCANLDAFTIKVKNENLDOLKNOIQEPCAKFA ADAT 545
Db 165 GLTASPGVAGATKQAKADHILQLCANDTCHIMSPKCYSQLMNNRPGCYV--COR 224
OY 546 REDPEKLELIMTRIQYQMSPM-SDFGTQPYEOMAIOMKKAQDGNRRCORCAEHL 604
Db 225 AADPGFDILIKLMNQIHQOLEMPDLKQEGTQMGYQVQVQLCKDAEALGDEQFYALHL 284
OY 605 KRYNEALQINDITRMIDATHTLETFYNEKKDKFAVIEDSDGDEYCDGJEDDOLK 664
Db 285 RRYNDALFTHDVTARADLMDLQDFYDRERTKTQMVAAES----- 325
OY 665 KRLKLEIDRELMTPFEENKMLKRLAENPEYENKLTKLRTIMEOYTRIEASAGLIF 724
Db 326 -----WLKLFDDKKNVLAESG-ENKLEKLELILKQF-GSHGHIGLIF 373
OY 725 TKTROSAYALSNQITENEKFAEYGVKAHHLIGAGHSEKPYMTONEQEVISKFRIGKIN 784
Db 374 TKTROSAYALSNQITENEKFAEYGVKAHHLIGAGHSEKPYMTONEQEVISKFRIGKIN 433
OY 785 LLIATVAESGIDIECNIVIRGLVTNEIAYOARGARADESTVILVAHSISVIEHE 844
Db 434 LLIATVAESGIDIECNIVIRGLVTNEIAYOARGARADESTVILVAHSISVIEHE 493
OY 845 TVNDPREKKMYKALHCVQMKPEEYVAKHILLOMOSIMEKKTKENIAKHYKNPSLIT 904
Db 494 LTVNEALEVMEKAAVAAYVQMDDEPRKAKRIQASLVKRAARAAREIQQQFLYEHVQ 553
OY 905 FLCKNSVLAESGEDIHYEKKHVMNTPFEKELTYVRENKALQKKCADYQNGETICK 962
Db 554 LVCINOMVAVGSGDLRKKEGHHVNVNPFVYTTTSGNPVYIKVKQMDPGIILGS 613
OY 963 KCGQAGMTVMYKSLDLPCLIRNFVYVFNKNSPKKQKKNVLPITFPNLDYSACCL 1019
Db 614 NCGEVAEFGMIVKSYVLPVLKIGS--MLETPRGKIQAKKMSRVFSPVFTLQNG 668

RESULT 8
OY 996C10 PRELIMINARY: PRT: 678 AA.
AC 096C10;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 76.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-COLON;
RA Strausberg R.;
RA Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC014949; AAH14949.1; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD.
DR Pfam: PF00271; Helicase_C; 1.
DR ATP-binding; Helicase; Hypothetical protein.
SQ SEQUENCE 678 AA: 76612 MW: 859E1749C7313D06 CRC64:

Query Match 25.5%; Score 1352; DB 4; Length 678;
Best Local Similarity 40.7%; Pred. No. 2.8e-66;
Matches 292; Conservative 136; Mismatches 237; Indels 52; Gaps 12;

```


OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sun Y.W.;
 RT "Rig-I, a human homolog gene of RNA helicase, is induced by retinoic acid during the differentiation of acute promyelocytic leukemia cell."
 RL Thesis (1997), Shanghai Institute of Hematology, Rui-Jin Hospital, Shanghai Second Medical University.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yi-Wu S.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF038963; AAD19826.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KM ATP-binding: Helicase.
 SO SEQUENCE 925 AA; 106613 MW; 5922BAF3DD0F00D4 CRC64;

Query Match 20.5%; Score 1086.5; DB 4; Length 925;
 Best Local Similarity 31.2%; Pred. No. 1,6e-51;
 Matches 325; Conservative 166; Mismatches 464; Indels 185; Gaps 35;

QY 12 RLILSCFRAVVKYIOVEPYLDLT--FLPAVEKQIORIVATSGMNOAVELLSTLEK 69
 DB 6 RSLQAFODYIRKTLPTLLTYLSTYAPWPFREEV-QYIAKKNKGPEALILFLKLE- 63
 QY 70 VMHIGMTREVEALRRTSGPLAARYMPELTDLPSPFENAH--DEYLOLILNLOPLVD 127
 DB 64 LOEGMGRGLADLDNAG--YSGLEYEISMD-----FKIKLELEYRLIKLQLEPK 116
 QY 128 KLVRLVDKCMEEELLTIEDNRRIAAENNGNESYRELLKRIYQ--K-NMFSAPLVNL 185
 DB 117 RIIPDTIIDP-LSECLINCECELLOICSTKGMAGAEKIVECLISDKIMPKTL- 171
 QY 186 RGTGNNELVQELGSDCESNAELENLSQDQPOVEQOLSTIVQRLKLVGMEN- 242
 DB 172 -----KLALEKERNKFSELMIVEKIGKVDTEDEKMTSDIQ-----IFYQELPEK 219
 QY 243 -NSSESFADSVYSESDTSLAGSVCLDESIGHNSNMQSDSGTMSD-SIENVAARAS 301
 DB 220 QNLSENGCPSEV---SDTL-----IS 239
 QY 302 PEPELOLRPYQMEVAQPALEGKNIICLPTSGKTRAVAYIAKDHLKKKASEPGKIV 361
 DB 240 P--FKFRNYQLDELAPAMKNTIICAPYCGCKTEFVSLICHHLL-KKILQGGKGVYF 295
 QY 362 LVKVLVLEQ--LFRKEPPLKKMYRVIGLSDTOLKISFEVAKSLIISTAFIL 418
 DB 296 FAHQDIPVEQKVSFSKYFE--RHGVRVIGISGATAEVNPVQIYENNIILITPQILV 352
 QY 419 NSLNLNENGADAGYOLDSFLIITDECHRTNKAFAVYNNIRHVLQWOLKNNRLKREKRV 476
 DB 353 N---NLKKGITP---SLISITLMTFDECHNTSKQHPYIMLNPENLIDOKLS--SS- 399
 QY 479 IPLPQILGLTASPGVAGATQAAAEHILKLANLAFTIKTYKENIDQIKNOIDEPCK 538
 DB 400 GPLPQVIGLTASVGVDAKMTDELDYICKLASLDASVATVAKHNELEIHOYVYKQVK 459
 QY 539 FALADATREDPEKE-----KLEIETRIOTYQCMSPKSNPTGTPKQW 581
 DB 460 FRKVESHSISKFYIIIAQLMRDTESLAKRICKLEMLNSLQIN-----FEPQIKLQW 512
 QY 582 AIOMEKKA---AKGKRKERV---AEHLKRYNEALQINDITPMIAVYHLETFYN 631
 DB 513 IIVYQKCMYQMPDKQDESRICKALFLYTSHLKRYDALIISFHAKMKALDYLAQTF 572
 QY 632 EEKDKKRAVIEDDSDEGDEYDQGDDEDEDLKKPLKLTDETRFLTFLENNKMKLKA 691

DB 573 NVRAAGFEEIED-----LTPREEMJDEHSVS 601
 QY 692 ENPEYENKLTLRNTIMROYTRTESARGIIFETRTROSVALSOWIENEX--FAFVG 749
 DB 602 RQPSNENKLEDCFLQDEYHINPPTIT-ILFVTRALVLDLKKMIESNFKSLAQS 660
 QY 750 KAHNHLGAGHSEFPKPMQNEQKEVYSKPR-TGKINLLIATVAEGDLKQINIVRYG 808
 DB 661 ----LTRGCKTQNTQNTMLTPAKCILDPAKAGDINIILATSVADQIDIAQVNLVLE 716
 QY 809 LVTNEIAMVOANGARADESTVYLVHAGSGVIEHETYNDPEKKMYKAIRVYMKPEE 868
 DB 717 YGVNFKIMQIOTGRGRARSKCEFLT-SNAGVIEKQIMYKEKKMNSILKQIMDAV 775
 QY 869 YAHKLTLEOMQSIEMKKMKTKRNIANKY--KNNSPLITFLCKNGSVLACSGHIVIEKM 926
 DB 776 FRKILHIQTH--EKFIQDSQEKPPVPDKENKKL--LCRKKCALACTYMAVAVIEE 829
 QY 927 HHVNMTPREKELIYVENKALQKQADYQINCEIIC--KCGQANGTMVHNGILDPCK 983
 DB 830 HTVLGDARKEGFVSRPHK-PKQSSFEKRAKIFCARQNGSHDQIHVKYIFELPYIK 888
 QY 984 IRNFVYVFKNNSTKKQYKRW 1003
 DB 889 IESFVEDIATGVOTLYSKW 908

RESULT 11
 ID 09GLV6 PRELIMINARY; PRT; 940 AA.
 AC 09GLV6;
 DT 01-MAR-2001 (TRENBLREL. 16, Created)
 DT 01-MAR-2001 (TRENBLREL. 16, last sequence update)
 DT 01-DEC-2001 (TRENBLREL. 19, last annotation update)
 DE RNA helicase.
 GN RHIV-1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCB1_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE:20261798; PubMed:10799277;
 RX Zhang X., Wang C., Schook L.B., Hawken R.J., Rutherford M.S.;
 RT "An RNA helicase, RHIV-1, induced by porcine reproductive and
 RT respiratory syndrome virus (PRRSV) is mapped on porcine chromosome
 RT 10q13."
 RL Microb. Pathog. 28:267-278(2000).
 DR EMBL: AF181119; AAG09428.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KM ATP-binding: Helicase.
 SO SEQUENCE 940 AA; 107583 MW; 118CA910B0AF7821 CRC64;

Query Match 19.2%; Score 1018; DB 6; Length 940;
 Best Local Similarity 30.1%; Pred. No. 9,6e-48;
 Matches 315; Conservative 158; Mismatches 369; Indels 204; Gaps 35;

QY 31 VLDYLT--FLPAVEKQIORIVATSGMNOAVELLSTLEKGMHIGMTREVEALRRTGS 88
 DB 25 ILSYAPWPFREDEV-OHIOENKKNKOPTPAASLFLOFLLE-LQEGMGRGLADLDNAG 81
 QY 89 PLARYMNPBLVD-LPSPFENAH--DEYLOLILNLOPLVDKLVYRDVLMKQME----- 140
 DB 82 -----YCG--LCAISISMPQKTEKLEYRSLRLQPEFKTTINPKDLEI IAH TLSQ 134
 QY 141 --ELLTIEDNRRIAAENNGNESGVRELLKRIYQENNFSAFLNVLROTGNNEI VOELT 198
 DB 135 ECEELIOLICSSKGLMA-----GAKEVCEILRS--DKENW----- 167

```

OY 199 GSDCSSENAELENISQVDPQVEBOLLSTVQPNLEKEVGMNNSPSSPADSSVSES 258
DB 168 -----PKTKLTALEKE-----ESRSELMWVKGAEVYK 196
OY 259 DTSIAGSGVSLDESLOCH-----NSNMGSDGTMGSDSDHENVAAARSPPELOLRP--Y 311
DB 197 MKLEDEDMKTCOVQIYKKEPEKONL-----NONGSSSAVPIYTPLEPRKY 244
OY 312 QMEVAOALGSKNIIITLPFGSGKTRVAVYIANDHLDKCKKASBPCKVIVLVKVLVEQ 371
DB 245 QLELALPAGKKNITICAPGCKCTEVSLLICEHNL--KKPRGRKGVVFPALQIVYRQ 303
OY 372 ---LFRKEFOPLKWVRVYLGSDGQDLKSPBEVYKSCIIISTANLNSLLENCE 428
DB 304 QKVSFSEHFE---RLGYKVGISGASISDVQVEQIVENSOTITLPOLANCLTN----- 395
OY 429 DAVV--OLSDPSLIIIDECHTNKEAVYNNIMHRYLMQKKNRLEKKNAPVLPQOIIQ 486
DB 356 ---GTIPSLVFTLMIPECHNTSKQHPYVIMSYIDRKIAGSS-----DSLQVQVIG 405
OY 487 LTVSPGVGATKQAKAEHLKLCANADAFITKVENLQDLKNOIOLPQKPAIADAIR 546
DB 406 LTVASVGVDAKNAEATEYICKLCASIDTSVATVRDNLLELPEVYKQKPKFRKVELKT 465
OY 547 EDPFKKLEIMTRIQYCO-----MSPHSDRTOPIYDQVATQMEKKA-- 590
DB 466 TDRKCTISOLMEIESLASIFEELQITTLGLFQIUNSNFQTOYEDQIVKQOEQVY 525
OY 591 ---KGNRKEKRC-----AEHLKRYNEALQINDTRMIPAVYHLEFVNEKDKKFAVI 641
DB 526 FQMPDKDKESRICKALFYSYSHIRYNDALINEMHAKMKALDYLDQFPFNIAAGF--- 582
OY 642 EDDSDGEGDEYCOGDEDEDLKKPLKLDTRFLMTLFFENNMKMKRLAENEVEYENKL 701
DB 583 -----DEIEQDL-----TOR-----FEEKLOELESISIDPSNENPKI 614
OY 702 TKLRNTIMEQYTRTESARGIIFTKTROSAVALSOMITENKFAVGVANHLIGCHSS 761
DB 615 RDLCTILOEYHINPPE--RTILVTKTRALVDALKMKIKINPKIS--FLKPSITLGGKTN 671
OY 762 EFPKPTONEQKEVYSKFRGKIN--LIATVAEGLDKICGNIVIRYGVITNEIAVQAR 820
DB 672 QNGMTPLPAGCVLDFTRDKDKKILITTSVADEGIDIAQCNVILLEYGVANTKMIQIR 731
OY 821 GRARADESTVYVAHSGSGVIEHETVNDPREKMYKAIHCVONMKDEYVAKHILELOKOS 880
DB 732 GGRGRAGSKCFLLT--ANADLIDKEKKNMYKEEMNGAILILQTDWBAVFKPKIHQIQ-- 788
OY 881 IMKKKKTKKNIAKHYKKNNSL-----TFLOKNCVIALGSGEDHIVFKMHHVMTPE 934
DB 789 -----REKILIRNOGKREPVPDKTKKLLCKCKKACAPCTADIRMEKCHFTVVGNA 840
OY 935 FKELYIVREKALOKKCADYQINGELIIC---KCGQAMGTMMVHKSDLQTKIRNFVYV 991
DB 841 FRERFVSKLHPK--PKSFGNIEKAKIYCARPDOSHDGILYVRIKAFEMR--KLESIVVD 899
OY 992 KNNSTKKQYKKNVELPITPPNDLYSE 1017
DB 900 IATGVQTVAHKMKDFNEFKISFDAE 925

```

```

RESULT 12
OY 909225 PRELIMINARY: PRT: 210 AA.
AC 090225;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 9130009C22RIK protein.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RC SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=CECUM;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
RA Atkawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Mikalido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombartts P.,
RA Morone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*
RT Nature 409:685-690(2001).
DR EMBL: AK018602; BAB31303.1;
DR MGD: MGI:1918836; 9130009C22RIK.
SQ
SEQUENCE 210 AA: 24441 MW; ZEBFDC06343933DF CRG64;
Query Match 18.1%; Score 960; DB 11; Length 210;
Best Local Similarity 86.7%; Pred. No. 1.9e-45;
Matches 182; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
OY 816 MYOARGARADASTVYLVHSGSGVIEHETVNDPREKMYKAIHCVONMKDEYVAKHILE 875
DB 1 MYOARGARADASTVYLVHSGSGVIEHETVNDPREKMYKAIHCVONMKDEYVAKHILE 60
OY 876 LQMSIMKKMKTKKNIAKHYKKNNSLITFLCKNCVIALGSGEDHIVIEKMIHVMTPPEF 935
DB 61 LQVSTLEKKMKKVERSIAQVNDNPSLITLLCKNCVIALGSGEDHIVIEKMIHVMTPPEF 120
OY 936 KELIYIVREKALOKKCADYQINGELIICGQAMGTMMVHKSDLQTKIRNFVYVFNNS 995
DB 121 KELIYIVREKALOKKCADYQINGELIICGQAMGTMMVHKSDLQTKIRNFVYVFNNS 180
OY 996 TKQYKKNVELPITPPNDLYSECLFSD 1025
DB 181 PKQYKKNVELPITPPNDLYSECLFSD 210

```

```

RESULT 13
OY 044165 PRELIMINARY: PRT: 1037 AA.
AC 044165;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 119.2 kDa protein.
GN F15H10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitida;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN (1)
RP STRAIN=BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RX None;
RT *Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.*;
RL Science 282:2012-2018(1998).
RN (2)

```

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2:
 RA Mollmann P., Murray J.:
 RT "The sequence of *C. elegans* cosmid F15B10."
 RN Submitted (Dec-1997) to the EMBL/Genbank/DBJ databases
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2:
 RA Waterston R.:
 RT "Direct Submission."
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases
 DR EMBL: AF036696; AAB88350.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR ATP-binding; Helicase; Hypothetical protein.
 KW SEQUENCE 1037 AA; 119188 MW; 63D189175DDA6776 CRC64;
 SQ

Query Match 12.9%; Score 687; DB 5; Length 1037;
 Best Local Similarity 25.0%; Pred. No. 1.7e-29;
 Matches 251; Conservative 169; Mismatches 408; Indels 174; Gaps 35;

OY 101 DLPSPEFAHADE--YLOLLNLOPTLVKLLVROLYDKCMEE----- 142
 DB 63 DLKSLNSAADDERLYKIDMYLQ--TYLCKTVHKLN--SNREVKLSDFEYILDHFFP 120
 OY 143 LTTIEDNRHIAAENNGNSGVRELKTRVQK---ENMFSAFLN-VLK-G--NNELVQ 195
 DB 121 LRFIEKRVAVLYDSYPOYIDAVALRKIEERNEEDNODSDFKLRLFPVLGGQAVY 180
 OY 196 EL--TSDCSSESAELENLSQVDGPOVEBQLSTVQPLKEKVMEN-----N 243
 DB 181 DMTYITSEKSSNNLDE-----AKQFLA-----KYLRLKRGELRFYQIIN 221
 OY 244 SSESFAADSVYSESDTSLAEGSVCLDSLGHS---NMGSDSGTMSGDSDE---ENV 296
 DB 222 ASRQRLNGRIYICPVHESATEMMVYGTALNTNRKRMNINRDNVQENRIPRLVIRSV 281
 OY 297 AARASEPELQLRPYOMEVAOPALEGKNITICLPITSGKFRVAVYIAKHLDKKKASP 356
 DB 282 RQRIHROQLCRNYQEEELCQVALQCKNTIVAPTSGKTVIAANTIKHEFSRSESKR 341
 OY 357 GKIVIVY-NKVLVLEQLFKREFQPLKMYRVLGLSGDQLKISFEVYKSCDIISTQ 415
 DB 342 FALFMTPRSMILNQ--AASISSYLDHYHTQIIOGSD--NVPINVLKSKDLIVAIPO 397
 OY 416 ILENSLLENLEGEADGV-----LSPFSLIIDECHTNKEAVYNNIMRYIMQKLNKR 470
 DB 398 MYNLCNEHNSLDDSRDLQDFLSTFTIIFPDECHNTYKNSPSYIMEY--HYKKNMG 455
 OY 471 LKKENKPVLPOLIGLTASPGVGATKQAKAEHILKLCANDATITIKIVENDOLAN 530
 DB 456 NMPEGH---SLPQIGLTASLTGDKNDQOVANYIAGLCASMDVADLSIVKDNLELKG 512
 OY 531 QIQEPCKRAIADATREDP---FKELLFETRIQ-----ITC 565
 DB 513 YSDIVDPKYLCCERSTDGIPGFTNRLTLMQVEGLIKETALRNEHIGIPRRQIETIR 572
 OY 566 QMSPMGDFTOP-----YEQWAIQMEKKAACKGRKEVVAEHLRKYNFALOINDIIR 618
 DB 573 DRPDSFIDPPADKHAQYQNVY-----NOMLVSGTSFRERQRTIILEAD 622
 OY 619 MI-DATYHLETFYNEKDKKFAVIEDSDSGDDEYCDDEDEDLKKPKLDETDKFLM 677
 DB 623 VLKECCCTLSYNINFRPVALNLTQDEM-----EY-----RTNFTIV 659
 OY 678 TLFFENKMLKRLAENPEYENKLTIRNTIMEQYRTESAGIIFTKIKOSAVALLSW 737
 DB 660 NMIRIRERHNOLVGTGSAENPMISKTVOYIVQONLORADS--RTIIFVIRIWEATILNKV 718

OY 738 ITENKEFAVGKAAHHLIGAGHSEPK---PMTQNEQEVSKPRTCKINLLIAIYAE 794
 DB 719 LNSNEELLMLGKSRMBSGLKSTASSADISAKQOKMEKLMADGDIRLIVASIAEE 778
 OY 795 GLDIKECNIVIRYGLVTEIAIMVQARGARADESTYVLVAHSGGVIEHETVNFREKMM 854
 DB 779 GLDVEGSLVIKYNATNEIAHVORGRGRALNSCEVLITNS-TALRQFESNNKAKSLM 837
 OY 855 YKAIQCVONMKE-----EYAKHLELOMOSIMEKKTKTRNIAKHKNPSILFLCK 908
 DB 838 SETISLNSPAEFKRCVDEESNKTPRILREDDTKAKIEQI-----NNIVAKILCK 892
 OY 909 NCSVLACGEDIVHEKMHVNMTPF-----KELYIVENKALQCKADYOINGILCK 963
 DB 893 KCEALICTSKDIR-SRNTQYLVCPGPFSLVAKTRILRDEQALIK-----YATGSLNCR 947
 OY 964 --CGQAMGTMMVHGLDPLCKIRNFVVFVFNNSKQYKKM 1003
 DB 948 ENCGILGLQLEVTNVDPCLSLALSTVLVEGTDRIIVKKM 989

RESULT 14
 OY 08VE79 PRELIMINARY; PRT; 143 AA.
 AC 08VE79;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Similar to RIKEN cDNA 9130009C22 gene (fragment).
 GN 9130009C22R1K.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.:
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC019605; AAB19605.1;
 DR MGD: MGI:1918836; 9130009C22R1K.
 FT NON-TER
 SQ SEQUENCE 143 AA; 16742 MW; 7F16E2CB4090AABA CRC64;

Query Match 12.8%; Score 678; DB 11; Length 143;
 Best Local Similarity 87.4%; Pred. No. 3.5e-30;
 Matches 125; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

OY 883 EKKMKTRNIAKHKNPSILITFLCKNCSVLACGEDIVHEKMHVNMTPFCKILYIVR 942
 DB 1 EKKMKVKSIAIQYNDNPSLITLLCKNCSMLVCSGENIHVEKMHVNMTPFCKILYIVR 60
 OY 943 ENKALQCKADYOINGELICKCGQAMGTMMVHKGDLDECLAIRNVVVFKNNSKQYKK 1002
 DB 61 ENKALQCKADYOINGELICKCGQAMGTMMVHKGDLDECLAIRNVVVFKNNSKQYKK 120
 OY 1003 WVELPPTPNLDYSECLFSD 1025
 DB 121 WVELPPTPNLDYSECLFSD 143

RESULT 15
 ID 017545 PRELIMINARY; PRT; 811 AA.
 AC 017545;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 93.4 kDa protein.
 GN COB10.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;

RN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2.
 RC MEDLINE-99069613; PubMed-9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Blanchard M., Bradshaw H.;
 RT "The sequence of *C. elegans* cosmid C01B10.";
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ database.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U58757; AAD47117.1; -.
 DR InterPro: IPR001410; DEAD.
 DR Pfam: PF00270; DEAD; 1.
 DR SMART: SM00487; HELICASE_C; 1.
 DR SMART: SM00490; HELIC_C; 1.
 KW ATP-binding; Helicase; Hypothetical protein.
 SQ SEQUENCE 811 AA; 93435 MW; 7PFAC1860EBF0A3 RC64;

Query Match 12.3%; Score 654; DB 5; Length 811;

Best local similarity 27.0%; Pred. No. 7,7e-28;

Matches 210; Conservative 147; Mismatches 318; Indels 102; Gaps 27;

QY 294 ENVAAPAPPELOLPYMEVAOPALEGKNIITLPTSGKRIEVAAYIAKDHLDDKKKA 353
 DB 49 ESYRORIHIOFOLCLRNVOELQVALQCKNITVAPTCGKIVIAANIKHEFEBSSE 108
 QY 354 SEGGKIVLV-NKVLVLEDFREKOPFLKKYRVIGLSDTQIKISFFVVKSCDIIIS 412
 DB 109 GKFKALFMTNPMNMLNDETRISS-----YLTOTIQSD-NVPTNNVIOSKDLIVA 159
 QY 413 TAOILENSLNLNGEDAGVO-----LSDSLIIDECHHINKEAVYNNIMRHLMCK 465
 DB 160 TPQMIVN-ICN-EHRPRLDDEYPPROFLSTIIIFEDCHNIVKNSPISNMYREY--HY 215
 QY 466 LKNNRLKKNKPVILPQILGLTASPGVGAATKQAKAEERILKCANLDAFTIKYKENTL 525
 DB 216 LKNNKMPREGH---SFPQIIGLTASIGTDKKNCMQVRSYIANGCANMDVKELSIVKDNL 272
 QY 526 DOLKNOIOEPCKFAIADATREDP-----FKELLETMTRIJYCOKSNMSDFTQD- 577
 DB 273 EELLDHNPFTVDVSCFSCNSDGPITEMFTKRLKQMOVEEDLITLKNETPKYEIPPT 332
 QY 578 -----YEOMATOMEKKAKKGNKREKVCACHLRKYNEALQINDITRMVAVTHLETYN 631
 DB 333 DKHNHYEMNINQNRGCVSLAGSRNKTLLIEVL-----ILKVOYHMSQMKIFWN 382
 QY 632 EKKDKKFAV--IEDSDGGDEYCDGDEDEDLKKPLKID-----LIDRPLMTLFF 681
 DB 383 YFRKRYKFKRIFOFODCFYALSYNINFPREVALAKYLEKELPERIRFTDN--MNLW 440
 QY 682 ENNKMLKRLAENPVEYENELTKLRNTIMQYRTESARGLITKTRQSAVALSOMITEN 741
 DB 441 DN--CHRELTVIGSAENPMIARTVOITLDONEOTSD-FALIVFRTKKKADFLNLYL--N 495
 QY 742 EKFAEYGAHHLIG---AGHSSEFPMTONEKKEVISKFTGKINLLIATVAEGID 798
 DB 496 DLHLHGLTSDMSGOKSTASADISAKOKOMKELKMFADRIHQILVSTVAEEGID 555
 QY 799 KECNIVIRYGLVTNETIAMVQARADESTVYLAHSGSGVIER--ETVNDPREKMYK 856
 DB 556 PCSLVIKYKYNATNETAHVQRGRARARNSKCVLITNS---IALHVQESNNIAKENIMTE 612

QY 857 AICHVQNNKPE-----EYAHKILELOMSIMEKKMKTKRNIAKHVKNPNPILFLCKNC 910
 DB 613 TSLIÖNSPGEFRQCVDEESNKKVMPRIQREDDKQRIKEQI-----NRNIYKIVCKKC 667
 QY 911 SYLACSGEDINHIEKMHVNMTPF-----KELYIVREKALQKCADYQINELIG---K 963
 DB 668 DTVLCTNKDIR-SKNIOYIVCNPGFWSLVRIPLPLEDORASNK---FNSTGSHCTGER 722
 QY 964 CGOAMCTMMVHKGDLPLPCIKIRNFYVFEKNNSTKKQYKKVVEL-----PIPPNLD 1014
 DB 723 CGSKIGQLIDVNTVNDPLCKVKSILLIESTNERILVQKMNILDEHPTPTIKGRD 779

Search completed: May 8, 2003, 15:54:11

Job time : 56 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:46:03 ; Search time 27 Seconds

(Without alignment)
1574.564 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311

Sequence: 1 MSNGYSTDENFRLISCFA.....LPITFVNDLYSCQLEFDED 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 402.5 | 7.6 | 1374 | YC9A_SCHPO | C09884 schizosach |
| 2 | 392.5 | 7.4 | 778 | YF05_MERJA | C58900 methanococ |
| 3 | 350 | 6.6 | 1845 | YM68_CAEEL | F45292 caenorhabdi |
| 4 | 303.5 | 5.7 | 993 | YIS2_YEAST | F40562 saccharomyc |
| 5 | 223 | 4.2 | 557 | YOH8_BACSU | F54509 bacillus su |
| 6 | 217.5 | 4.1 | 2663 | CENE_HUMAN | C02224 homo sapien |
| 7 | 214.5 | 4.0 | 1790 | USO1_YEAST | F25366 saccharomyc |
| 8 | 204.5 | 3.9 | 2163 | BRR2_YEAST | F32639 saccharomyc |
| 9 | 197.5 | 3.7 | 1163 | SBCC_GLOAB | C97481 clostridium |
| 10 | 197 | 3.7 | 2230 | GOG4_HUMAN | C13439 homo sapien |
| 11 | 193.5 | 3.6 | 663 | UVB8_STRAA | C09417 staphylococ |
| 12 | 193.5 | 3.6 | 2017 | MYSN_DROME | C99323 drosophila |
| 13 | 193 | 3.6 | 715 | HEIS_PYRAB | C00089 pyrococcus |
| 14 | 191 | 3.6 | 3911 | AKA9_HUMAN | C06966 h-a-kinase |
| 15 | 189.5 | 3.5 | 3210 | CENE_HUMAN | F49454 homo sapien |
| 16 | 188.5 | 3.5 | 1433 | REST_CHICK | C42184 gallus gall |
| 17 | 187.5 | 3.5 | 720 | HEIS_PYRAB | C73946 pyrococcus |
| 18 | 183.5 | 3.5 | 663 | UVB8_PUSNN | C09417 fusobacteri |
| 19 | 183.5 | 3.5 | 1875 | MLP1_YEAST | F12452 saccharomyc |
| 20 | 182.5 | 3.4 | 3660 | DMD_CHICK | F11533 gallus gall |
| 21 | 181.5 | 3.4 | 2869 | RBPI_PLAVB | C00798 plasmodium |
| 22 | 180.5 | 3.4 | 479 | DBPA_BACSU | F42305 bacillus su |
| 23 | 180.5 | 3.4 | 976 | SCPI_HUMAN | C15431 homo sapien |
| 24 | 180.5 | 3.4 | 1131 | YAB9_YEAST | F81360 saccharomyc |
| 25 | 180 | 3.4 | 2704 | BPA1_HUMAN | C03001 homo sapien |
| 26 | 179.5 | 3.4 | 607 | DB10_NICSY | F46942 nicotiana s |
| 27 | 179.5 | 3.4 | 1938 | MYH6_MOUSE | C02566 mus muscula |
| 28 | 179.5 | 3.4 | 2022 | ANT1_ONCVO | F21249 onchocerca |
| 29 | 179 | 3.4 | 564 | ROK1_YEAST | F45818 saccharomyc |
| 30 | 179 | 3.4 | 662 | UVB8_STREN | C54966 streptococ |
| 31 | 179 | 3.4 | 754 | YAJ3_SCHPO | C09903 schizosach |
| 32 | 179 | 3.4 | 2748 | NUM1_YEAST | C00402 saccharomyc |
| 33 | 178.5 | 3.4 | 1539 | Y373_HUMAN | C15078 homo sapien |

| | | | | | | |
|----|-------|-----|------|---|-------------|--------------------|
| 34 | 177.5 | 3.3 | 444 | 1 | SRMB_ECOLI | P21507 escherichia |
| 35 | 177 | 3.3 | 1679 | 1 | Y109_YEAST | P41457 saccharomyc |
| 36 | 176.5 | 3.3 | 2116 | 1 | MY52_DICDI | P04791 dictyosteli |
| 37 | 176 | 3.3 | 658 | 1 | UVB8_LISTIN | O94784 listeria in |
| 38 | 174 | 3.3 | 687 | 1 | DB73_DROME | P24680 drosophila |
| 39 | 173.5 | 3.3 | 1727 | 1 | ALM1_SCHPO | O94018 schizosach |
| 40 | 173 | 3.3 | 637 | 1 | NTP1_FOMPV | O72907 fowlpox vir |
| 41 | 173 | 3.3 | 2349 | 1 | TPR_HUMAN | P12273 homo sapien |
| 42 | 170 | 3.2 | 1102 | 1 | MYSC_CHICK | P20016 gallus gall |
| 43 | 169 | 3.2 | 658 | 1 | UVB8_LISMO | P20016 gallus gall |
| 44 | 168.5 | 3.2 | 646 | 1 | UVB8_METHA | O24542 methanobact |
| 45 | 168.5 | 3.2 | 748 | 1 | CLPE_LACIA | O94109 lactococcus |

ALIGNMENTS

| RESULT 1 | YC9A_SCHPO | STANDARD: | PRT: | 1374 AA. |
|----------|--|-----------|------|----------|
| ID | YC9A_SCHPO | | | |
| AC | C09884 | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | | |
| DE | Putative helicase C584.10C (EC 3.6.1.-). | | | |
| GN | SPPC584.10C. | | | |
| OS | Schizosaccharomyces pombe (fission yeast). | | | |
| OC | Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; | | | |
| OC | Schizosaccharomycetales; Schizosaccharomycetaceae; | | | |
| OX | NCBI_TaxID=4896; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=972; | | | |
| RC | MEDLINE=21848401; PubMed=11859360; | | | |
| RA | Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., | | | |
| RA | Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S., | | | |
| RA | Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., | | | |
| RA | Collins M., Connor R., Cronin A., Davis P., Felkewell J., Fraser A., | | | |
| RA | Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodge G., | | | |
| RA | Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jarvis K., | | | |
| RA | James K., Jones L., Jones M., Leather S., McDonald S., McLean J., | | | |
| RA | Mooney P., Moule S., Mungall K., Murphy L., Nibbellett D., Nisbell J., | | | |
| RA | Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E., | | | |
| RA | Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., | | | |
| RA | Skellton J., Simmonds S., Squares R., Squares S., Stevens K., | | | |
| RA | Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., | | | |
| RA | Woodward J., Volckstreit G., Aert R., Robben J., Grimprey P., | | | |
| RA | Wetjens I., Vanstreels E., Rieger M., Schaefer M., Muelier A.M., | | | |
| RA | Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hillert H., | | | |
| RA | Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl J.M., | | | |
| RA | Bier P., Zimmermann W., Medler H., Wandt R., Furelle P., | | | |
| RA | Galbault A., Gadieu E., Dreano S., Gloux S., Leleux V., Muller S., | | | |
| RA | Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., | | | |
| RA | Lucas M., Rochet M., Gallardin C., Tallada V.A., Garcon A., Jude G., | | | |
| RA | Dagmar R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., | | | |
| RA | Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsting S.L., | | | |
| RA | Cerrutti L., Lowe F., McCombie W.R., Faulsen I., Potashkin I., | | | |
| RA | Shpakovski G.V., Ussery D., Barrett B.G., Nurse P., | | | |
| RT | The genome sequence of Schizosaccharomyces pombe. | | | |
| RL | Nature 415:871-880(2002). | | | |
| CC | - SIMILARITY: BELONGS TO THE HELICASE FAMILY. | | | |
| CC | - SIMILARITY: CONTAINS 2 RNASE III DOMAINS. | | | |
| CC | - This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL database. | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| DR | EMBL; AL032824; CAB37423.1; - | | | |


```

OY 426 NGEADG-VOLSDPSLIIIDECHTNKAAVYNNIMRHYLMKLNKLNK:HKPVLIPOL 484
DB 123 ----AGRINDEFILLADBAHHTTGDHAY-----AFVAKKFKDK-----CHT 161
OY 485 LGITASPGVGGATKQAKAEHILKLCANDAFITKTVKNLNDOLK-----529
DB 162 LGITASPG-----SDIDKVMCEINIGIEHVEVTRHDEIVKQYIAKVKLIPRID 213
OY 530 -NOIOECKKFAIADATRE-----DPFKLEIMRIOTVCOMSWSOP 573
DB 214 PNEFKALK--LINEALKERLIKLDAGVINSIADYKTELIDLNKLP-----262
OY 574 GTQPYEQMATQMEKKAAGNKRKERYCAEHLRKYNLOINDITRMIDA-----1THLD 627
DB 263 -----DEEVYELIKVCEHLKMAKKELLES:GKSVFVNTYN 300
OY 628 TFYNEKDKKFAVIEDSDGDEYCDGDEDEDLKKPLKDETDREFLLEFENNKKML 687
DB 301 KLSMORTKSAKSIVNDE-----KVRBAVNLKKSQVHEHPKL- 336
OY 688 KRLAENPEYENKLTIRNTIMEQYTRTESAGCIIFTKROSAYALSON: TENEFAEY 747
DB 337 -----GKVDVNVKNILK-----NKDERIIIFADYRDVVERKIVNI 375
OY 748 GKVAHHLIGAGHSSEFPKPTONOEKEYISFKRTGKINLIATTAVERGL: KECNIVIRY 807
DB 376 GIRAIRITGA-KKEGKMSQKQEDIEATEPKK-EGSVIVTSVSEBGI: PBYVNIIF 433
OY 808 GLVTNELIAYQANGARADE--STVYLVAHSGSVLEHDI:VNDPREKMA:NAIHCVQNM- 864
DB 434 EAPVSEIRFIORGRARAGCGKVVYLIAGTADAYRSL:YKREMKKL- 489
OY 865 -----KPEYAKHILLOMSIMEKMKTKRINAKIKY-----NINSLITFLCKN 909
DB 490 YLLNKLQKFEESKEKEETEELKEKELESKTAVKEFKEEKTKXVITLDFIKO 549
OY 910 CSV--LACSGED-----IHVIEKHHVNMTP-----EFK-1-----YVRE 943
DB 550 IEVKEKSEFEDKIKOPIKIPKRIKITYVREKNNAKL:HNANINIELK:ILEVDGYLSD 609
OY 944 NKALOKKADYQINGEI 960
DB 610 RVYVERKTAEDFVNSII 626

RESULT 3
ID YMG8_CAEEL STANDARD: PRT; 1845 AA.
AC P34529;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical helicas K12H4.8 in chromosome III.
GN K12H4.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhaditida; Rhaditoidae;
OC Rhaditidae; Pezodermidae; Caenorhabditis.
OX NCBI_TaxID:6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jiet M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaulian K.,
RA Waterston R., Watson A., Wellstock L., Wilkinson-Spoat J.,

```

```

RA Wohlsman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
RN (2)
RP REVISIONS.
RA Waterston R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 RNASE III DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: L14331; AAA28101.2;
CC PIR: S44849; S44849.
CC WormRep: K12H4.8; CE25057.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR001159; DS_RBD.
CC InterPro: IPR005034; DUF283.
CC InterPro: IPR001650; Helicase_C.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR000999; RNase_3.
CC Pfam: PF00270; DEAD_1.
CC Pfam: PF00271; Helicase_C_1.
CC Pfam: PF00636; Ribonuclease_3; 2.
CC Pfam: PF02170; PAZ; 1.
CC Pfam: PF03368; DUF283; 1.
CC SMART: SM00487; DEXDC; 1.
CC SMART: SM00358; DSRM; 1.
CC SMART: SM00490; HELICC; 1.
CC SMART: SM00535; RIBOC; 2.
CC PROSITE: PS50137; DS_RBD; 1.
CC PROSITE: PS50821; PAZ; 1.
CC PROSITE: PS00517; RNase_3; 1; 1.
CC PROSITE: PS50142; RNase_3; 2; 2.
CC KW Hypothetical protein; Helicase; ATP-binding; Hydrolase; Nuclease;
KW Endonuclease; Repeat.
KW NP_BIND 33
KW FT DOMAIN 759 915 ATP (POTENTIAL).
KW FT SITE 145 148 PAZ.
KW FT DOMAIN 1316 1524 DECH BOX.
KW FT DOMAIN 1578 1740 RNASE III 1.
KW FT DOMAIN 1768 1831 RNASE III 2.
KW FT DRBM.
SQ SEQUENCE 1845 AA; 210922 MW; 4A96EA2922FE1D9A CRC64;

Query Match
Best Local Similarity 23.5%; Score 350; DB 1; Length 1845;
Matches 175; Conservative 118; Mismatches 249; Indels 2; Gaps 37;

OY 309 RPYQMEVAPALEGKNIIICLPTGSGKTRVAYIANDH-----LDKRRKASE:NAVIV 361
DB 14 RDYQVELDKATK-KNTIVOLGTGSGKTFIALLKEYGQVLPAPIDQCK-KAF 67
OY 362 LVKKVLLVQGLFKKEKQFLKKRYVIGLSGD-----QKISFPEVAVSCII:ISAD 416
DB 68 VKEKVMVLEQ--QAIHIEVHSFKVGVNGQTSGLMSKKECDGFMKRIHVIVITAO 124
OY 417 LENSLLNLENGEDAGVOLFSLIIIDECHT-NKEAVYNNIMRHYLMKLNKLNK:KREN 475
DB 125 L-----LDIRIRAYLIKEDKCYLIPDECHNALGSHPRISMVY-----KIKKD 170
OY 476 KVPVLPOLIGLTPASPGVGGATKQAKAEHILKLCANDAFITKTVKNLNDOLK:IDEP 535
DB 171 KPV---PVLIGLTPAS--IAKAVPERKLMQKLKLESAMDS-VLETFASD-LVS:SVI:GAMP 224

```

OY 536 -----OKKFAIA--DAREDPREKLELEIMTRIQTICQWSPNNDTQ-----576
 DB 225 YEYVILCKDEIGLIGIPNEDYIEIDFVAVNITTEPHPIDDUPREHIDSKITR 284
 OY 577 -----PYEOMALOMKKAOKKONKREYCAELHLYNALOIND-TIMNID-AV-HLE 627
 DB 285 AVFROLOPNAAM-----RFAOVWEKELGKIISQVLPDKTLFIMAKTISMT 331
 OY 628 TFFNEKDKKFAVIEDSDGDEGDEYCDGDEDEDLKKPLKDEIDREFLMTFFENKML 687
 DB 332 TI-----KRLT-----EPENKKIKSIALRPYIVICVIRILFEIL 365
 OY 688 KRLAENPEYENKTKIRNTIMQYTTESAGIIFTKROSAYAL-----SOWITE 740
 DB 366 ETP--NEFOKERN-KLEKA-----EHLSAIIFVDKRYATAYSLILMRHISW---410
 OY 741 NEKFAEYGVAAHLIGAG-----HSSEKPMYONEOKKEVISKPRCKNLIATTVAEGL 796
 DB 411 EPMFK--FVMPDYVYGASGNILASSDQGL-HKROIFEVLRKHFRNEINCLIAISVLEEGV 467
 OY 797 DIKECNIVIRYGLVTNEIAMVOAGRADESTYVLAHSG-----SGVIE 842
 DB 468 DVKOCNLYIKFDRPLDMRSYVQSKGRARRAGSRVILVEHKOTAAVCSKLIISDIPTRLVP 527
 OY 843 HEVYNDRKEMKKAHICVONMMP--EYAHKILELOMOSIMKKKKTKNIAKHKNP 900
 DB 528 HNOIPIEENGVTK--YCAELLPIINSPIKHA1-----VLKNPMPNK-----569
 OY 901 SLATFLCKNCVLAISGE-DIHVIEK-----MHVNNTPPE-----FRE 917
 DB 570 AQMAVVALEACROHLEBEIDNLLPKGRESIAKLEHIDEPDEYAPGIAAKVSKKKNO 629
 OY 938 LYIVREKAL-----OKKADY 954
 DB 630 LYDKIARALNESFVEADKECFIY 653
 RESULT 4
 YIS2_YEAST
 ID YIS2_YEAST STANDARD; PRT; 993 AA.
 AC P40562;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative ATP-dependent RNA helicase YIR002C.
 GN YIR002C OR YIR2C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Hornslett T.S., Hunt S., Jagsels K., Jones M.,
 RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
 RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,
 RA Schwager C., Zimmermann J., Sander C., Ansorge W.;
 RT "Nucleotide sequence and analysis of the centromeric region of yeast
 RT chromosome IX";
 RL Yeast 11:61-76(1995).
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
 CC SUBFAMILY.
 CC

CC This SWISS-PROT entry is copyright. It is produced through collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Consortium -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 DR EMBL: Z38062; CAA86204.1;
 DR EMBL: X79743; -; NOT_ANNOTATED_CDS.
 DR PIR: S48436; S48436.
 DR SGD: S0001441; YIR002C.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_Box.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD_1.
 DR Pfam: PF00271; Helicase_C_1.
 DR SMART: SM00487; DEXDC_1.
 DR SMART: SM00490; HELIC_1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE; PAUSE_NEG.
 KW Hypothetical protein; ATP-binding; RNA-binding; Helicase.
 FT NP_BIND 107 114 ATP (POTENTIAL).
 FT SITE 209 212 DEAD_BOX.
 SQ SEQUENCE 993 AA: 114057 MW: 47400950.543171F CRC64;
 Query Match 5.7%; Score 303.5; DB 1; Length 993;
 Best Local Similarity 20.3%; Pred. No. 3.8e-08;
 Matches 143; Conservative 118; Mismatches 246; Indels 14; Gaps 25;
 OY 307 QLRPYOMEVAPALAEKNNIIIGLPGSGCTRAVYIADHLDKKKASPKCVIIVKV 366
 DB 86 EVADYQYTVHKS-L-FONTLCALPTGCKTFIATVYMLVFRMTKA---KILFATPR 140
 OY 367 LVEQLFRKEQFLKMYRVLGSGD-----TOLKISFPEVYKSCDIISTAOILENS 420
 DB 141 PLVAGQ-----IKACGIGTIPSDQFALLIDSKRKREIMANKRVFFATPVEND 192
 OY 421 LNLKNGEDAGVQLSDFSIIITDECHHTNKEAVYNNIMHNYIMOKLKNKPKVYP 480
 DB 193 L-----KRGVLDPRDYLVCVIDEAMRATGSSAYTNVVF-----IIRFNSSV--234
 OY 481 LPOILGLTASPGVGATGAKAEHILCANLAFATIK-TYKENDLOKKNLOPCKKF 539
 DB 235 --RLALATITP-----ASDLGVQEVYNNLDISKIEIRTEESMDIVYKKKKKKKI 284
 OY 540 AIAADATREDPEKLELEIMTRIQTICQWSPNNDTQ-----EKKAA 590
 DB 285 EV-----PLLEIIDIIE-----QIGMAVKPVLQDAIEIGIYEHTPSQUNAF 327
 OY 591 KKGKNERKCA-----EHLRKNEAL-----QINDTIMDAVTHLETFYNEKKAAY 640
 DB 328 KAMQOSOKTIANPTIPEGIKMNFPILOLLNNVGMLEKRLKIY-GIRTFNYPUNK-----382
 OY 641 IEDSDGDEGDEYCDGDEDEDLKKPLKDEIDREFLMTFFEN-NKMLKRIAFN--FE 695
 DB 383 -----CTEFTTKYMLK-----STNKIAAEFTYIPILKINKNQNTATLSFK 424
 OY 696 YENE-KLRLKNTIMEQYTTESAGIIFTKROSAYALSGMI-----738
 DB 425 FVGHCKLQCVDELDMPQKRGSDSRVILFELRSALIEYFIDSVADDOIRPHILIGQ 484
 OY 739 -----TENKFAEYG-----VAAIHILJAGH 759
 DB 485 ARAKGFDEVKYTRKHAPKGRKKVERLHRQDEKLEAEKTRANDKLER-SAKRTSSSE 544
 OY 760 SEEFPMYONEOKKEVISKPRCKNLIATTVAEGLDEKCNIVIRYGLVINFIAMVVA 819
 DB 545 EAOJGSMOKMKEVYIHNFKKEVAVLVCTSIGEGCLDIGEVDLTICTDTSSIKNIQR 604
 OY 820 RGR-ARADESTVLAHSGSGVIEHETVNDFERKMYRAHCVONMK-----FF- 867
 DB 605 MGRTRKRRGKIVLFFSSNESYKFERAMEDYSTIOLALISKQIDVKKSDRIIPHIITFG 664

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 mitosis." Nature 359:536-539(1992).
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Throner D.A., Jordan M.A., Schaar B.T., Yen T.J., Mills L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 microtubule motor." Embo J. 14:918-926(1995).
 RL Embo J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 interactions with the kinetochore proteins CENP-F and HUBB1." J.
 Cell Biol. 143:49-63(1998).
 RL J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND HUBB1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CONGSSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE. AND IS
 QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).
 CC
 DR EMBL: Z15005; CAA78727.1;
 DR PIR: S28261; S28261.
 DR HSSP: P17119; 3KAR.
 DR Gene: HGNC:1856; CENPE.
 DR MIM: 117143;
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam: PF00225; kinesin_1.
 DR PRINTS: PRO0380; KINESINHEAVY.
 DR SMART: SM00129; KISC. 1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS00667; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 SO SEQUENCE 2663 AA; CEF13880C8-8C88 CK54;
 Query Match 4.1%; Score 217.5; DR 1; Length 2663;
 Best local Similarity 18.6%; Pred. No. 0.0074;
 Matches 209; Conservative 185; Mismatches 421; Indels 307; Gaps 47;
 OY 9 ENFRYLSCFRARVKKYIOVEPVLDYLFPAVKKQIOYVAFSGNMVAVELLSTLK 68
 DB 1040 EQGRKIFSLQENKELQOMLESYIAEKQDKTDKKNITMTLE---NQRHKLGLGFLKK 1096
 OY 69 GVMHLCMTREYV---EALRTGSPPLAARVKNPDLTDSPSFENAHDTQLQNLMLQIP 123
 DB 1097 -----QOEIVAOEKNAIKKEGELSRGCDRLAEVEEKIKKESQDQKQQLLVNCEH 1149

OY 124 TLVDKLLVNDV--LDKCMEEELTLE--DRNR1AAEN--NGNDSVRELK_RIVVNT-- 175
 DB 1150 MSEMKKINEIENKELNKKELTLEHMETERLELAOKLNENYEFVKSTIKKKYKLIQ 1209
 OY 176 -----NMESAFVNLRTG-----NNELVOELTSGSCSNMPTNL 212
 DB 1210 KSETERDHLRGVIRETEVTLQTKBELKIANHILKEHETIDELRRS-VSKRVQDIIAT 1268
 OY 213 SVDGPOV-----EEDLLSTTVQNLKEVWGMNNSSEFSADSVASNDIS 261
 DB 1269 QDLKSHTKLOEIRPVLHDEGELL-----PVKKVSETGETNNLELLELQSTIKNS-IT 1322
 OY 262 LAEGVSCDSESGHNSNGSGTSGSDSDENNAARSPPELOLRVQKRVQVAFATE 321
 DB 1323 LARTEM---ERLRNKKF-----QESOEIKSLTKERDNIKTIKKALVKKHPLT- 1368
 OY 322 GKNIIICLPTSGKTRVAVYIAK-----DHLDKKAKSEPGKVIVLVKAVIIVYCL 372
 DB 1369 -----KEHIRETLAKIQESOSKQOSLMMKKEDNETKI-----VPPMPS- 1408
 OY 373 FRKEPPLKKMYRVIGLSGTQAKISPE---VYKSCDIIISTAOILENSIANFNC-- 427
 DB 1409 FPKQDALLRIETEMIGLS--KRLQSDHDKMSVAKENDLQRLQEVLOSSEIXG;KENIK 1466
 OY 428 EDAGVQLSPSLIIDECHHTKKEAVYNNIMRHYLMOKIKNNRLKKENKPVIPVIVIGL 487
 DB 1467 EIVAKHLETEEBELVAACHCKLKEDEETIN-----ELRVNLSEKETE----- 1506
 OY 488 TASPQVGGATQOKA-----EEHILKCANLDAFTTKYKENIDOLKNQICVYKKF 539
 DB 1507 -----ISTIOKOLEAINDLQKIOEIVKEKQOLNIKOISEVQEVANEIK-GKKIKLKK- 1558
 OY 540 AIADATREDPFEKLEIMTRIQTVCQ--MSPMSDGTQPYQOMAIOMKKVAKKQNR-- 595
 DB 1559 --AKDSALOSIESKMLTNRLOESOEIOIKKEEKMKRVQALQIEDQCKNTKEI 1616
 OY 596 -----KERVQ-AEHLKRYNMLQIN-----DTIRVIVV-AYT 624
 DB 1617 VAKMSESOEYQFLKMAVNETOEMCEIEILKKQFQETKNDNITENITKLTQLIHE 1676
 OY 625 HLETFYNEKKKFAVIEDSDGDDDEDCDDDEDDIKKPLKLD----- 670
 DB 1677 NLEEMSVYKER-----DLRSVEETLKVENDQKLENIRITTDLEKQELKIVIMHL 1730
 OY 671 ----ETDRFLMLTFEENKMLKRLAENPEYENKL-----TKLRNTIMVYVTE 716
 DB 1731 KEHOETIDKLRGIVSEKTEINISNMOKDLEHSDALKADQKIOELRIAHMH KQVFTI 1790
 OY 717 ESARGIIFTKROSAYALSQWITEN-----EKAQEVGAHNLIGAGHS-----SFFK 764
 DB 1791 DKLRGIVSEKTKLSMOKOLENSNAKLOEKIOELKANEHOLITLKKDVNETQKVVSTME 1850
 OY 765 PMTQ--NEQKEVYSFPRGKINILATTVAAEEDGIKECNIVIRGLVTNEIAVQVAGGR 822
 DB 1851 QLKQIKQDSLSLKLEIENLNL--AQELHEHLEEMK-----SVKKIKEDN 1893
 OY 823 ARADESTYLVLAHSSGYI-----EH-ETVNDPFRKKMYKAIH 859
 DB 1894 LRRVEETIKLERDQKESLOETKARDLEIQOELKTARMLSKENKETEYDLKRIKTIQ 1953
 OY 860 GVOYNNK-----PEYVAHKLLEIOMOSIM-----EKKMKTKRIAHYNNKNSLI 903
 DB 1954 ISDIQKDLKSDKDELQKIOELQKKELOLRLVKKEDVNNHKKINEMEOELKKQFQIN --- 2009
 OY 904 TELCKNCSYLAGSGEDIVIEKMHVNTTPEPKELIYRENNK 945
 DB 2010 -YLCK-----CEMDNPOLTKLHE-----SLEIRIVAKER 2039
 RESULT 7
 USOL_YEAST
 ID_USOL_YEAST
 AC P25386;
 STANDARD: PRT: 1790 AA.

DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein US01.
 GN US01 OR INT1 OR YD058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NX NCBI_TaxID:4932.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-X2180-1A;
 RX MEDLINE-91185402; PubMed-2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, us01, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae."
 RL J. Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.; to the EMBL/GenBank/DBJ databases
 RL Submitted (FEB-1993)
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERNS CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X54378; CA38253.1; -;
 DR EMBL: L03188; AAB00143.1; -;
 DR EMBL: U53668; AAB6659.1; -;
 DR PIR: A38455; A38455.
 DR SGD: S0002216; US01.
 DR InterPro: IPR002017; Spectrin.
 KM Transport: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDER (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B215E9F4818 CMC64.
 Query Match 4.08; Score 214.5; Dh 1; Length 1790;
 Best Local Similarity 19.38; Pred.No. 0.0027;
 Matches 180; Conservative 158; Mismatches 36; Indels 259; Gaps 37;
 38 LPAEVKEDQIORTVATSGNMAVELLSTLEKGVNHLGWTAEFVALRRTS----- 88

Db 904 LKEDIAAKITEIKAINENLEEMKIQCNLSKEKEHI--SKELVEYKSRFSQHNLVAKUT 961
 QY 89 -----PLAARY--ANPELTOLPSPFFENADDEVILQQLNLQPLTVOKL-----LVREV 134
 Db 962 EKLSLANNKDKQAENESLKAVERSKNNSIQLSNL--QNKIDMSQELNFOIIFGS 1019
 QY 135 LDKMEDELLTIDRRIRIAAENNGNSGVRELKRIYQKEMFSFLVLRK-----T 188
 Db 1020 IENITDQLKTTISDL-----EOTKEELISKSDSKDEYSQILKFKLETATT 1068
 QY 189 GNNELVQELTSGDSESNAEIENLSQVDPQVEEQLLSTTVQPNLEKEWGNENSS 248
 Db 1069 ANEENNKI--SLTTRTELE-----AELAAKNLKNEL-----ETRLFTS 1108
 QY 249 FADSVSVESDTSVLAQSVSLDESIGHNSMGSQSGTWS--DSDEENAAKASPEPELO 307
 Db 1109 EKALKEVKEENEHLKEEKLQLEKEATEKQQLNSLANESLEKEHEDLA-----Q 1160
 QY 308 LRPQMEVQAPALEGNIIICLP--TSGKTRVAYYIAKDHLDKKKKASEVQXVILVN 364
 Db 1161 LKRYEDQIAKERYNEDELSQLNDEITSTQENESIKKKNDEGEVAKMS----- 1212
 QY 365 KVLVLEQLFRKEFQPLKKRYVIGLSGDTOLKISEPEVAKSCDIIITAOIENSLNL 424
 Db 1213 -----TSEQSNLKK-----SEIDALNLQIKELKKNE--TNEASTLE-----SI 1250
 QY 425 ENGEDAGVOLDSFLIITDECHHTNKRAYNNIMRYLMOKLNN-----RLKLNK 476
 Db 1251 KVSSEVETVQIKELQ--DECNKEKEV-----SELDEKILKASEDKSKSYLQENISE 1299
 QY 477 PV-----IPPLQILGTASPGVCGATKAKAKKEHLKLCANLDAFTIKYKEN 524
 Db 1300 KIEELDAKTTELKIQLEKTLNLSK-----AKKESSELSRLKTKSSGER--KNATQ 1350
 QY 525 LDOLKNOIQEPCKFAIADATREDPEKELLEITRIQTYCOMSPNSDGTQPEYEW 584
 Db 1351 LELKLNELQIKNOAFE-----KERKL-----LNBSSTITQESKJNT 1389
 QY 585 MEKKAAGKGRKRVCAEHLKRYNEALQINDTIRMDATHTLETFTNEKDKKFAVIED 643
 Db 1390 LEDELRLIQENEL-----KAKEIDNTSELEKSLSNDELLEKONTIKSNLDE 1439
 QY 644 -----SDSGDGDGYC-----DGDDEDLKKPLKLDOTRFLMTLFFENKKY--KHLAEMP 694
 Db 1440 ILSTKQITRNDKLLSIEDNKRDLSEKQELAAQESKAY-----EENLKLKEES 1493
 QY 695 -----EYENKILTKLNTIMQYTRTESARGIIFTRROSAYALASOMIENEFVAV 747
 Db 1494 SKKAELEKSEKEMKKLESTIESNEITELKSSMTI-----RKDEKLEQ-----SKSAEE 1544
 QY 748 GVAHAHLIGASHSEKRPMTQNEQKYEISKFTGKINLLATTVAEGLDKIKCNVIR- 806
 Db 1545 DIK-----NLQHEKSDLSIR-----INSESDIDIEIKSKLRI 1576
 QY 807 -----YGLVTMEIAMVQARGARADESTYVLAHSGSVIEHEIVNDFRKKMKAH 859
 Db 1577 EAKSGSELETVKQELNNAOKIRINAEENT-----VLKSLDEIDELIKQJAE 1625
 QY 860 CVONMKPEEYAH--KILELOMOSIMEKKKATK 889
 Db 1626 IKSNOEKELTIRLKLDELSDTOOKAKXSE 1658
 RESULT 8
 BRR2 YEAST
 ID BRR2 YEAST STANDARD; PRT; 2163 AA.
 AC P32639;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pre-mRNA splicing helicase BRR2 (EC 3.6.1.-) (Protein Smu24b).
 GN BRR2 OR RSS1 OR SMU246 OR YER172C OR SYCP-ORF66.

OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 Aviles E., Berno A., Brennan T., Carpenter J., Chang E., Cherry J.M.,
 Chung E., Duncan M., Guzman A., Hartzell G., Hunkeler-Smith S.,
 Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 Patel F.X., Roberts D., Sehl P., Schramm S., Shougen T., Smith V.,
 Taylor P., Wei Y., Yelton M., Botstein D., Davis K.M.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-169 FROM N.A.
 RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
 Wei Y., Taylor P., Nakahara K., Roberts D., Davis K.M.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RC STRAIN=DBY473;
 RC MEDLINE=96304576; PubMed=8722763;
 RA Noble S.M., Guthrie C.;
 RT Identification of novel genes required for yeast pre-mRNA splicing
 RT by means of cold-sensitive mutations.;
 RL Genetics 143:67-80(1996).
 RL [4]
 RP CHARACTERIZATION.
 RC MEDLINE=96324408; PubMed=8670905;
 RA Lauber J., Fabrizio P., Teigelkamp S., Lane W.S., Hartmann E.,
 Luehmann R.;
 RT The HeLa 200 kDa U5 snRNP-specific protein and its homologue in
 RT Saccharomyces cerevisiae are members of the UEXH-box protein family of
 RT putative RNA helicases.;
 RL EMBO J. 15:4001-4015(1996).
 CC -1- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potencia.).
 CC -1- DOMAIN: COMPOSED OF TWO SIMILAR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SIMILARITY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U18922; AAB64699.1;
 DR PIR: S30856; S30856.
 DR SGD: S0000974; BRR2.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004179; Sec63.
 DR Pfam: PF00270; DEAD_2.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; Sec63; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KM Hydrolase; Helicase; mRNA processing; mRNA splicing; Spliceosome;
 KM Nuclear protein; ATP-binding; Repeat.
 KM DOMAIN 521 923
 FT DOMAIN 1370 1757
 FT NP_BIND 521 528 ATP (POTENTIAL).
 FT NP_BIND 1370 1377 ATP (POTENTIAL).
 FT SITE 634 637 DEH BOX.
 FT SITE 1474 1477 DDAH BOX.
 SQ SEQUENCE 2163 AA; 246183 MW; DFAFEE387168D944 CRC64;
 Query Match 3.9%; Score 204.5; DR 1; Length 2163;
 Best Local Similarity 19.1%; Pred. No. 0.011;

Matches 162; Conservative 136; Mismatches 247; Indels 40; Gaps 40;
 QY 127 DKLVRDVLDMKME-----ELLTIED-----RNR--IAAAE 156
 DB 1305 DTSVIODLSEKILINDIETLEHNPALEQVLQVLLDFENISLAFLTKNRSTLFWELPLAK 364
 QY 157 NNGNESGVELLRKYOKENMFSAFLNVLROTGNNELVOELTSGSCSSESNATLENSVD 216
 DB 365 STENE--IPNLEKVKAK-----GLNDLVEQYKFRRTTSKRILD--SGSD 406
 QY 217 GQVVE-----QLSTVQNLKEYWGMENNESSSEFADS--VSESPTSLAVGSVSTLD 271
 DB 407 QPSSSAKRTKSNRAIPVIDLE-----KIKPDESKNTLVTKSLVQDQPK--- 454
 QY 272 ESLGHSNNGSDSGTMSDSDSENVAAASPEPEQLR----- 309
 DB 455 -----RYKKPYDEIHIPASPKPIVDLEKEITSLPDMCEAFISTSTSLN 500
 QY 310 PYQMEVAQPALEG-KNIITCLPTGSGKTRVAVY-----IAKHLDKKKKASIPKVIYV 363
 DB 501 PLSQKVFHAAPFGDSNMLCAPTSGKTNIALTVLAKALSHYNNPKTKLNISAKIYVI 560
 QY 364 N--KVLVEQLPRKEFO--PELKKWRYVIGSGDTOLKISPEVWKGCDIITSTVQILE 418
 DB 561 APLKALVQPOV--REQRRLAFL--GIKVAELTGSRLS--RKQIDETQVIVSTIFKMD 613
 QY 419 NSLLNLNENGADAGVQSLIDFSLIIDECH--HTNKEAVYNNIM--RHYLMOKLKNLTKKEN 475
 DB 614 ITRNSNN--LAIVEL--VRLIIDEIHLLHDQRPVLESIVAFPMASKYQV---- 663
 QY 476 KQVIVPLPOLGITASGVGATKQAKAEHILKCANIDAFITIKYKENTLQKMLQLEP 535
 DB 664 -----YPRILIGLSAT-----LPNVEDGRF-LRVPKGLLEFFDSNRP 700
 QY 536 CKKFAIADTRDPPEKLEIMTRIQYQMSPSDGTQPEQWALOMEKKAKKGNR 595
 DB 701 C-----PLS-----QPPGKIKENSIKNIKAKA 721
 QY 596 KERVQAEH-LRRYNALQINDTIRMIDAVTILETFYNEKKKFAVIEDDSIEGDMDEYC 654
 DB 722 MNDACEYKELFESINQNL--IVFHSRKET----- 750
 QY 655 DQDEDEDLKKLKLDEIDRFIMTLTFPENNKKLKLAEPEYENKTLKRLNTIMVYTR 714
 DB 751 -----SKRTATLKNKFAEN-----ITIKTLTK 772
 QY 715 TESRAGILFTKTRQSAVALSQWITENEKFAEVGAKAHHLIGAGHSSEFKPTQWVQKEV 774
 DB 773 NNGSKQILTK--EAANVLDPSL--KRLIESIGTGH--AG-----ITRSNGSIS 816
 QY 775 ISKFRGKINLLIATVAEGLDIKECNLVIIRGLV-----TNFIANVYVARG 821
 DB 817 EDLFDAGILQVLACTATLAMGVNLPANHVIYLIKQVYSPKGSWQLSPQVLAQWILHAG 876
 QY 822 RARADESTYLVLAHGSQVIEHETVNDREKKMYAALNCYQNMKKEEVAHKLILQMSI 881
 DB 877 RPRYD--TF-----GEGII-----ITDSNVQYVLSVNLQDPIESQPSKIVINLNAEV 924
 QY 882 MEKKMKTKRN 891
 DB 925 VAGNIKCRND 934
 RESULT 9
 SBCC_CLOAB
 ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
 AC O97FK1;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nuclease sbccD subunit C.
 GN SBCC OR CAC2736.
 OS Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1486:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RA MEDLINE-21359325; PubMed-1146286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F.V., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum*;
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -1- FUNCTION: sbcd cleaves DNA hairpin structures. These structures
 CC can inhibit DNA replication and are intermediates in certain DNA
 CC recombination reactions. The complex acts as a 3'-5' double
 CC strand exonuclease that can open hairpins. It also has a 5'
 CC single-strand endonuclease activity (By similarity).
 CC -1- SUBUNIT: Heterodimer of sbcd and sbcd (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SMC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE007771; AAK0682.1; -
 DR InterPro: IPR00439; ABC_transpor.
 KM Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
 KM DNA recombination; ATP-binding; Coiled coil; Complete proteome.
 FT NP_BIND 35 42 ATP (POTENTIAL).
 FT DOMAIN 197 415 COILED COIL (POTENTIAL).
 FT FT 446 1003 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1163 AA; 135507 MW; CEF0BD221507A92 CRC64;
 Query Match 3.78; Score 197.5; DB 1; Length 1163;
 Best Local Similarity 18.88; Pred. No. 0.011;
 Matches 197; Conservative 187; Mismatches 399; Indels 263; Gaps 45;
 QY 31 VLDYLTLPFAVEKEQIQRTVATSGNMQAVELLSTLEKGVHMGTRFEVALRIGSP 89
 DB 44 ILDSITLSYGEVARKSSNPNNTNCSLNSVFEQISGKELKRLVREF---RKNKNT 99
 QY 90 LAARYNPDELTDLPSPFENAHDEYQLNLLOPTLVKLLVSDVLDCKMPEELLITEDR 149
 DB 100 GSVRSKSAKIVDIGDEVEYLEG-----AKSVNKKCCTIGLSLIDDF 142
 QY 150 NRIAAEN-----NGNSSVRELLKRIYQKEM---FSALINVIKQ-TGNNELV 194
 DB 143 TRIVVLPQKSEFLKLEGRER--RNMLERFLNLOEYGDLSFLKARKIKEREKENVLV 200
 QY 195 QELTG-----SDCSSENAEINLSQVQDQVEQLSTVOPNLEKEVMENNESSSF 249
 DB 201 GELGTYENINDVAKERRELLKNNDFNENSKREYLAKEEYNKREVMQIETIEKNR 260
 QY 250 ADSSVSESSTSLAEGSVCLDESIGHNSNNGSDSGTMSDDEFNVAATASPEPEQLR 309
 DB 261 VRKDLMEKKD-----EIDLKERRAFIG-ESSSKVK 289
 QY 310 PY-----QMEVAPALEGKNIITCLPTGSKTRVAVYIADNLDKRRKASEPQVY 360
 DB 290 PYIDNVENTLKQIDILKEQLISRE-----NTMKALISLEKFMCKKSIANKKREK 339
 QY 361 VL-----VANKVLVQLRFRKEFQPLK-KMYRVIGLSGDTOLKIFPPEV---VKSDDI 410
 DB 340 ALPKFMKIHNIIDAIKEKQLDINKIKELKRLQKIEKISLEANSKKEELLKONIKDIDL 399
 QY 411 ISTAQILENSLNLNENGEDAGVQLSD--FSLIIDE-CHHTN----- 449

DB 400 TLKIONLESKIDNLKVPPEEYKKNINIEGIFLNRNYDEKLKHKKNKIGLDCQPOVLEKAS 459
 QY 450 -KNAVNNIMRHILMOKLKNRNLKKNKRVYLPOLIGLTASGVGA-TKQAK-PHHL 507
 DB 460 KKEMLFKLEERSKIDYTKRLQDLNKKD-FPKDVL-LTPOKLDNSQKMAKISEYE 517
 QY 508 KLCANLDAF-----TIKTVENIDOLKNQIO-----EPCCKPAIDATRE----- 547
 DB 518 SLKASLRVENSQVLRTRKEEMTKLEDKISKVNIKIESLETENMAHVLREKKEEACP 577
 QY 548 -----DPFEKLELINTRIQTYCOMSPNSDF-GTQPYE-----QMA-TKKAA 590
 DB 578 VCGSVHIRKGFREVDLKALETLK-----SELGFEKKRFEEMETVMCEASIKVEKNI 632
 QY 591 KKN-----KREVCACHELRKYN-EALQINDIR-MINAVHLETF 629
 DB 633 KKLNESINNLGEEFKEVLSLESMKKFNLYEKVKNFLKIDLDNIKOLSEKNIYEVE 692
 QY 630 YNEEK-----DKPFAVIEDSDGDDDEYCDGDEDEDLKPRLKDETDFLM:LFPEN 683
 DB 693 YQKEKVEYKQCEKRYIDKLESEEE-AIKFENEVATITENKALKIAD-----FKTEM 744
 QY 684 NKMLKRLAENPEYENKELTKRLNTIMEQYTRTESARGIIFTKTROSAYALSOW:INENK 743
 DB 745 KEILKERVAVAEAGE-IKDLRLMLNIRHTEKQO-----LMDKCSRLKEI-KKN- 793
 QY 744 FAEVYKAHHLIAGSHSEKPRPTONQOKEVISKFRGKINLLIATVAEFGIT-KKCN 803
 DB 794 -AEL-----KKDKIINEKIELI-KNVGVGLNDLYELKEKIEG-TIKK-- 834
 QY 804 VIRGLVTNEIAMVQARGRADESTYVLVAHSGSVIEHEETVNDREKMMYKAHVCN 863
 DB 835 EEOYNLDCKKMNIEDKYKCSDE---IKYHSNLSLDRKND-----IKLKNK 882
 QY 864 MKPEYAKHLELDOMOSIMEKKMKTKRNIAKHKNPSLITLCKNCSVLACS3-IHVI 923
 DB 883 ILMEKFEINIEKAKENVLNDKEINLLKSDVEKYNELSKV-----NGAVEVLS- 930
 QY 924 EKMHHVMPEFKELIVRNKALQK 949
 DB 931 KKLKRNKLT---EKKWLEIONNREK 953
 RESULT 10
 G064_HUMAN STANDARD: PRT: 2230 AA.
 AC Q13439; Q14436; Q13270; Q13654;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Golgi autoantigen, golgin subfamily A 4 (Trans-golgi p230) (56 kDa
 DE golgin) (Golgin-245) (72.1 protein).
 GN GOLG4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RA MEDLINE-96215236; PubMed-8626529;
 RA Eilich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.H.;
 RT Molecular characterization of trans-Golgi p230: a human periphera
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif*;
 RT J. Biol. Chem. 271:8328-8337(1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Seelig H.P.;
 RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE=Placenta;
 MEDLINE-96125112; PubMed-8537393;

RA Fritzel M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.H.,
 RT "Molecular characterization of goldfin-245, a novel Golgi complex
 RT protein containing a granin signature."
 RT J. Biol. Chem. 270:31262-31268(1995).
 RN (4)
 RP SEQUENCE OF 524-672 FROM N.A.
 RC TISSUE-Gastric fundus;
 RA Balague C.;
 RL Thesis (1994), Instituto municipal de investigation medica, Spain.
 CC -1- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
 CC GOLGI.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOSOLIC FACE OF THE
 CC GOLGI MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms: 1 (shown here), 2, 3
 CC and 4; are produced by alternative splicing.
 CC -1- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SCOGREN'S SYNDROME AND
 CC IN HEPATITIS B.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: UA1740; AAC50434.1;
 DR EMBL: X82834; CAAS8041.1;
 DR EMBL: U31906; AAC51791.1;
 DR EMBL: X76942; CAAS4261.1;
 DR GenBank: H0NC:4427; GOLGAA.
 DR MIM: 602509;
 DR InterPro: IPR000237; GRIP_domain.
 DR Pfam: PF01465; GRIP_1.
 KM Golgi stack; Antigen; Coiled coil; Alternative splicing.
 FT DOMAIN 133 237
 FT DOMAIN 276 1011
 FT DOMAIN 1033 1214
 FT DOMAIN 1259 2152
 FT VARSPLIC 2154 2185
 FT
 FT VARSPLIC 2186 2230
 FT VARSPLIC 2103 2109
 FT VARSPLIC 2222 2230
 FT CONFLICT 188 188
 FT CONFLICT 220 220
 FT CONFLICT 276 276
 FT CONFLICT 584 584
 FT CONFLICT 628 628
 FT CONFLICT 630 630
 FT CONFLICT 682 682
 SQ SEQUENCE 2230 AA: 261139 MW: 388733 DBIENAB6134 CRCAI
 Query Match 3.7%; Score 197; DB 1; Length 230;
 Best Local Similarity 17.9%; Pred. No. 0.028;
 Matches 185; Conservative 212; Mismatches 366; Indels 276; Gaps 46;
 OY 15 ISCFRRVKNYIOVEPVLDYTFLLPAEVKQIOIVATSGNMQVALLTEKQWMLG 74
 DB 384 IAOILRSRIRK-----MTQGELEBREKQEKSERA--FELEKALSTIAOK----- 425
 OY 75 WRFPEVALRRGSPPLAAYMNPDLTLPSPFENAHDEYLLNLQPIIVDKLLKRV 134
 DB 426 -TEE-----ARRKLAEM-DEQIKTEKTESE--EKISIQDEI SHVKQEVVIV 469
 OY 135 LDKMEELTLIEDRRIRIAAENNGSVRELKRIYQKMMWSEAFVLVHQTGNELV 194
 DB 470 MKKSSEQALQKQKHEKELARE-----QELTKLQTRKREPODMVLEKSSQSYL 523
 OY 195 QELTSGDCSESAEINISQVDPQVEQLLSTVQPNLEKQWENNSSSESTFASV 254
 DB 524 K-----ISOKEQOESLA-LELELQKKAILL-----ISNKLRLDQ 560

OY 255 VESDSTSLAEGVSCIDSELAGHNSNGSDSGT-MSDSEENVAAASGFE---PEIQLRP 310
 DB 561 EAEYTRILLESLEKSLKQENKNSKDLAVHLEAKKHNKKEITVWEKHEKTELSLK 620
 OY 311 YQNEVAPALECKNNIIICPTSGKTRAVVYIAKHDLKKKASPGKIVLVNVL-- 368
 DB 621 HQ-----QDAL-----WTEKLOVLKQOYQYEMEKREKCEQETLKKR--IFQ 665
 OY 369 -VEQLFRKPEQFLKRWYVIGLSDGTQKISPEVWASCDIISTAQIDENSLINEN 426
 DB 666 AHLEENNEKTEKLDVQKQELSLSE-----LSVYKA-----RHKLEELSYIKD 712
 OY 427 GEDAGVQSDPSLIIDECHTNKEAVYNNIRHMLQKLNKRNKREKVPYLPVLIG 486
 DB 713 QTDKMQEELAM--DEQKNHQQOVDSIIKEH-----EVSIOETREKALKQDINOELL- 764
 OY 487 LTASPGVQATQAAKEEHLKCANLDAFTIKYKENLDQKNOIOERCKKFAI--A 542
 DB 765 -----LKERDKLKEHOAH-----VENLEADIKREGELQOASAKLDVFSYQ 807
 OY 543 DATRED--PPEKLEIMTRIOTYCOMSPSPDGTQPYEQMAIOEKKAAKGNKRLVVC 600
 DB 808 SATHQTKAYEQDLAQLOOK-----LADLETE-----RILTKQAYEVAQKALVC 853
 OY 601 AE--HLKRYNALQINDTIRMIDAYTHLETFYNEEKQKRFVVIDSDSDGDDDEV--D 657
 DB 854 TELDAH-----KIQVDLMQOLEK-----QNSEMQKVSILTOVESKLE-----DGN 896
 OY 658 EEDDLKKPLKLDERTDRPLMTLFFENKKMLKRLAE-----NPEYE----- 697
 DB 897 KEOBOTKOLV--EKENMLQHRGOKKEIETLTQKLSAKEDSIHILNEEYTFKRNK 954
 OY 698 -----NEKLTKLNTIMEQYTRTEESARGIIFTKTROSAYALQWITENEKFAEYGVKAH 752
 DB 955 KMEYKQAKKEMQDELKLLKLDQEAKK--KELENALALESO--KEKQFNAKMIEMA 1007
 OY 753 HLGAGHSEFPW--TQNEQKEVISKFRGTQINILIA-----TTVAEGLDILK--NI- 803
 DB 1008 QANSAGISDAVSRLETKQEQTESLFEVHRRELNDVVISIWEKKLNOQAEELQETIHEIDLO 1067
 OY 804 -----VIRYGLTNEITAMVQARGRADESYVYLVASGCVIHEHETNPRE 851
 DB 1068 EKEQVAAELKQKILFEGCKEEM-----AKETITWL-----KEEVKQOTTINILQ 1113
 OY 852 KMVYKAIHCVQNMKPEE--YAH-KILE-----LOMOSIMEKKKTRRNIAKHKNPS 901
 DB 1114 QLKQSAH-VNSLAODETKLKAHLEKLEVDLMSKLENTPIQOVELKMLAEDEKRYVS 1172
 OY 902 LITFLCKNSVLAAGGEDIHVEKHHVNMTPFEKELYIVENKALQKCAIYO----- 955
 DB 1173 ELT-----SKLTDEEFQSLKSSHE-----KSNKSLKESLFFKRLSEEL 1213
 OY 956 -INGELICKQGA 967
 DB 1214 AIDLDICKKTEA 1226
 RESULT 11
 UVRB_STAM
 ID UVRB_STAM STANDARD: PRT: 663 AA.
 AC Q99YL7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Excituclease ABC subunit B.
 UVRB OR SAV0758 OR SA0713.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 NCBI_TaxID=158878, 158879;
 RN [1]
 RP SEQUENCE FROM N.A.

DR InterPro: IPR000048; IQ-region.
 DR InterPro: IPR004008; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head. 1.
 DR Pfam: PF00612; IQ. 1.
 DR Pfam: PF01576; Myosin_tail. 1.
 DR Pfam: PF02736; Myosin_N. 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head. 1.
 DR SMART: SM00015; IQ. 1.
 DR SMART: SM00242; MYSC. 1.
 DR PROSITE: PS50096; IQ. 1.
 KW MYOSIN: Alternative splicing: Coiled coil: Actin-binding:
 KW ATP-binding: Calmodulin-binding.
 FT DOMAIN 1 829 MYOSIN HEAD-LIKE.
 FT DOMAIN 830 859 IQ.
 FT DOMAIN 886 2017 COILED COIL (POTENTIAL).
 FT NF_BIND 225 232 ATP.
 FT DOMAIN 250 260 25 KDA/50 KDA JUNCTION.
 FT DOMAIN 682 694 50 KDA/20 KDA JUNCTION.
 FT DOMAIN 705 727 ACTIN-BINDING.
 FT DOMAIN 742 758 REACTIVE SULFHYDRYL/ACTIN-BINDING.
 FT DOMAIN 1303 2017 LIGHT MEMBRANE (1MM).
 FT DOMAIN 1303 1970 ALPHA-HELICAL TAILPIECE (1MM).
 FT DOMAIN 1971 2017 GLOBULAR TAILPIECE.
 FT VARSLIC 1 45 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 2017 AA: 232016 MW: 736302BAH+2528 CRC64:

Query Match 3.6%; Score 193.5; DH 1; Length 2017;
 Best local similarity 18.3%; Pred. No. 0.037;
 Matches 206; Conservative 182; Mismatches 386; Idents 351; Gaps 45;

23 KMTQVEPVLDYITFLPAEYKEIOITVATSGMMAVELLSTLEKGVHLLGNTREFVEA 82
 880 RLYTKVKKPL-----EVTKEKLVOKEDELKOVREKIDTLAKNTQ--YERKYQA 929

83 L-----RRTGSPLAARYNNPELTLP-----SSEFNAHDEYLT--- 117

930 LVKRTLAEDLOAEIELCAEESRSRLMAR--KOELEIMOELETRIEERVALG 987

118 -----LNL-----LOPTLVDKLLVAVD-----KMEPELLTEDS NRITAAENNN 159

988 EKKLELNIQDLEEOLEEEBAHQKQOLEKVQDAKIKKEEDLALTDQ,NQKLLKKEKK 1047

160 NEEGVELLKRIYOKENWES--AFVLNLTQGTNNELVQEL-----TGSNSESNAET 209

1048 LEEBRANDLSQTLAEEBEKKAKHLAKAKHEATITTELEERLHKQOCORPSHMKR 1107

210 -----ENLSOVDPQPV-----EEOILLST----- 227

1108 VADLKQLNERRVQVDEMQAQLAKKEBELTQTLIRIDHSATKATQAKA,RELESQIAET 1167

228 -----TVQPLEKEVGMEN-----NSSSESPADSSVVSSEI--SLAEQSVSG 269

1168 QEDLEAKKARAKAEKVRRLDSELEALKNELDISLDTTAAQELRSKRFGLATIKKSL 1227

270 LDESLHNSNMG-----SDSGTMSQSDS--ENVAARASPEPELQRPYQWVAQPALEGN 324

1228 EEEETVHNEGVLDMRKHSHSELNSINDOLEN-----LRKATVLEKKAKGTILFAEN 1277

325 IILCLPTGSGKTRVAVYIAKHDKKKASPEKVIIVLNKYLIVQULFKKFFOPPLAKW 384

1278 ADLATELRS-----VNSSQENDRRKKQAE--SQIAELQVKKIAELERA--KSEIQ----- 1323

385 YRVIGLSGDTQLIKISPEVYKSCDIIISTAQIENSLMLNMGHAGV-----QLSD 436

1324 -----EKCTKIQQAENITNQLAEELKASAAVVASNMESQLTE 1363

437 FSLITLIDEGH-----HTNKEAVYNNIMRRHYLMQKLNKLRKRPVPLPQI 484

1364 AQQLEEFTRQKLGSSKLRQIESKEALQEOLEED---DEAKRNERKLAETVTTQMOET 1420

QY 485 LGITSPGVGATKQAKAEHILKLCANLDAFTIKTVKENLQNLQIO--EYTKK+1A 542

DB 1421 -----KKKAE-----DAULAKELEGGKRLKLNKDLEALIK+VKH1A 1457

QY 543 DATREDFEKKEL---LEITRIQITQCMSPMSDFGQPTPEQMAIOMERKKAANKKRI 599

DB 1458 QNDRLKSKKKKIQSELEDAI--LELEAKRTVLELEKKONKPKILAEBAKLSQIAQIRD 1516

QY 600 CAEHLKRYNEALQINDTIRIMADYTHLETFYNEEDKKFAVIEDSDSEGDDEY+IXCHD 659

DB 1517 TEREAREKETVYLSRSLDAPDKIEDLEKKR-----TQNLDELQALQ---GTAD 1566

QY 660 EDDKLKPLLDLDFRLMTLFFENKMKRLAEENPEYNE-KLT---KIR-----NTIME 710

DB 1569 ---KNVHELEKAKRAL-----ESQIAELKQNELEDDQITLTDKALRLLEVNMQALRS 1618

QY 711 QYTR-----TEESARGII-----PRTROSAAVLSQWITENKIAAGCVK 750

DB 1619 QEEHRLAKEEGAEERKRLVQRLDLETEDEERKORTAAVASKKKEGDKI+ETTME 1678

QY 751 ANHLI---GAGHSSEKPTQN-----DQKEYISKFRGKINLILATIVAEQG 795

DB 1679 MNHKVEDALKHAKKLOAQVKALDADAEAKAKAEIQLAKSEADQKVALFAVLQ--- 1735

QY 796 LDKECNVIRGLVTNEIAMYQARGARADSTVLYAHSGSVIEHETVNDREKKMY 855

DB 1736 -----LTEDLASSE--RARRAAETEREDELAEELIANNANKSLMIDRKRLA 1780

QY 856 KAIHCQVNNKPEEYAHKIL-----ELQMSIMEKKKTKNIKAKHKNNSI+ITLAK 908

DB 1781 RALTEELLEEERQSSSEVLDDAAAROLOIETTELANENSNSQK--ENGANILR--- 1835

QY 909 MCSVLACSGEDIHVIEKMHVNMTPFEKELYIVRENKALQKKCADY 955

DB 1836 -----RONKELKALAEIE 1844

RESULT 13

HELS_PYRAB STANDARD: PRT: 715 AA.

AC Q9V0A9;

DT 16-OCT-2001 (Rel. 40; created)

DT 16-OCT-2001 (Rel. 40; last sequence update)

DT 16-OCT-2001 (Rel. 40; last annotation update)

DE Putative sk12-type helicase (EC 3.6.1.1).

GN PAB0592.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus

OX NCBI_TaxID=29292;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=GES / Orsay;

RA Helig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -1. SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.1st.sib.ch/annotation/](http://www.1st.sib.ch/annotation/or send an email to license@sib.ch) or send an email to license@sib.ch).

CC

CC

DR EMBL: AJ248285; CAB49795.1; -

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR003583; HNH.1.

DR InterPro: IPR001650; Helicase-C.

DR InterPro: IPR000445; HNH.

DR Pfam: PF00270; DEAD; 1.

| QY | 307 | QLRPYQME-VAPQPLLEGKNIILICLTPTSGSKTRVVA-VYIANKDHLQKKKKA-IPGKIYIVLVN | 364 |
|----|-----|--|-----|
| Db | 23 | ELYPPQALATLSGVLEGNNLLVLAIPFASGKTLIAISIALA-----NKLFIYSGKAVYIVP | 76 |
| QY | 365 | KVLVEOLFRRKEFOPLKMYRVIGLSGDTDLKISFPPEVVKSCDIITISIACTLFNSLNL | 424 |
| Db | 77 | LKALAEKFR-EFKDMELGLKVMATGDDYOSK---DEMLGKYDIIATPFRKFDL--- | 129 |
| QY | 425 | ENGEDAGVQLDSFSLIIIDECG---HTNKAAVYNNIMRYLIMQKLNRRKKENKRVIVPL | 481 |
| Db | 130 | ---RHGSSMIDVKKLVIDEIHLIGSRDQATLEFIIYQIMGR----- | 169 |
| QY | 482 | POIIIGLTASPEVGCGATKAKAAEHLIKLCANIDATITIKVENLDQKNL--DEPCKKFAI | 541 |
| Db | 170 | AOIIIGLSATIG-----NLELKAIVN----- | 190 |
| QY | 542 | ADATRPEDFEKKLEIMTRIQTQYCSMSPSDPGTQPYEQWALUYMEKKAKKGNRRERYCA | 601 |
| Db | 191 | AKLIRSDMRPKL-----NRG----- | 206 |
| QY | 602 | EHLKRYNEALQINDTRIMIDAVTILETPEYNEEKQKFAVIEDDSDEGDHYGCDDEDED | 661 |
| Db | 207 | -----VFY-----OGFVWED-----GGSFRYSNSEE--- | 228 |
| QY | 662 | DLKKPKLIDETDRPLMLTFEENKKMLKLAEMREVENEKLYTKARTIMLVTRTESASAG | 721 |
| Db | 229 | -----LYIDAVKKKGALVAVN---NRKKAETALELAK---VKKVNLDBFELR----- | 271 |
| QY | 722 | IIFKTRQSAVALSOWITENENEKFAEV---GVKANHILGAGHSSEKPRMLNEQKEVYSKF | 778 |
| Db | 272 | -----ELRRLASLEENPT-NKTLAKALLSGVAFNH---AG-----LQSPERVLEVDMF | 316 |
| QY | 779 | RTGKINLIATTVAAEGDLKECNIVIR-----YGLVTNLIAMVQ---ARGRAD- | 826 |
| Db | 317 | RKGLKVVAVAPRTLSAGINTPAFRYIIRDITWRYSFEGMERIIVLEVOQAMGRAGRPVDE | 376 |
| QY | 827 | ---ESTYIVLVANGSGVIEHETVNDPRKEMMYKALINQVONMKREEVANKIILEQMSIMEX | 884 |
| Db | 377 | VGEALIVSTTEPSPVLVDHYIKG-----KPEKL---PQUSNSESILRS | 416 |
| QY | 885 | KMKTKRNIAKH-YKNKPSLITFLCKNCASVLAQSGEDHVI-FKMHVNNHIFEFKELYIVR | 942 |
| Db | 417 | QILAL-IATGYSEFKELIYDFLER--TFYANQGGDPRMLEKIRI-----IYFLD | 464 |
| QY | 943 | ENKALQKKCA 953 | |
| Db | 465 | ENEFLEVTLED 475 | |

DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (Hyperion protein) (Vollrat protein) (Centrosome- and golgi-localized
DE PKN-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIA0803.
ON Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 4).
RP TISSUE-Brain;
RC MEDLINE=98151389; PubMed=9482789;
RX Lin J.W., Myszyński M., Madhavan R., Sealock R., Kim J.U., Sheng M.:
RA "Vollrat, a novel protein of neuromuscular junction and brain that
RT interacts with specific splice variants of NMDA receptor subunit
RT NR1.";
RT J. Neurosci. 18:2017-2027(1998).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM 2). AND VARIANT GLN-1347 INS.
RP MEDLINE=99219864; PubMed=10202149;
RX Wlczak O., Skalneq B.S., Keryer G., Bornens M., Tasken K.,
RA Jahsen T., Oststavi S.:
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RT protein located in the centrosome, AKAP450.";
RL EMBO J. 18:1858-1868(1999).
RN [3]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RP TISSUE-Brain;
RC MEDLINE=99287934; PubMed=10358086;
RX Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H.,
RA no Y.:
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
RT anchors multiple signaling enzymes to centrosome and the actin
RT apparatus.";
RT J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP Kemmer W.A., Deiss S., Schwarz U.:
RA "Cloning of Hyperion.";
RT Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
RN [5]
RN SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RP TISSUE-gastric parietal cell;
RC MEDLINE=99115654; PubMed=9915845;
RX Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.S.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.:
RT "AKAP50, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.";
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RN SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RP TISSUE-Lymphoblast;
RC Hinds K., Sauterer C., Becker M., Hawkins N.:
RA Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RN [7]
RN SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RP TISSUE-Lung;
RC Milgram S.L., Goldenring J.R., Schmidt P.H.:
RT "AKAP50: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [8]
RN SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RP TISSUE-Brain;
RC MEDLINE=99087487; PubMed=9872452;
RX Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.:
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [9]
RN SEQUENCE OF 17-1800 FROM N.A.
RP Wu X., Graves T., Bradshaw H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNIT OF PROTEIN KINASE
 CC A. SCARFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
 CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
 CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
 CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
 CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
 CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
 CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
 CC -1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
 CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
 CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC Epsilon.
 CC -1- SUBCELLULAR LOCATION: CYTOSOL IN MANY CELL TYPES AND
 CC CYTOPLASMIC IN PARIENTAL CELLS.
 CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 3/CG-
 CC NAP, 4/YOTIAO, 5 AND 6/MAP350; ARE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
 CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
 CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
 CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
 CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.
 CC -1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
 CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AJ131693; CAB40713.1; -
 DR EMBL: AB019691; BAA78718.1; -
 DR EMBL: AJ010770; CAA09361.1; -
 DR EMBL: AF026245; AAB86384.1; -
 DR EMBL: AF083037; AAD22767.1; -
 DR EMBL: AC004013; AAB96867.1; -
 DR EMBL: AF091711; AAD39719.1; -
 DR EMBL: AB018346; BAA34523.1; -
 DR EMBL: AC000066; AAC60380.1; -
 DR Gene: HGNC:379; AKAP9.
 DR MIM: 604001; -
 KW Coiled coil; Alternative splicing; Polymorphism.
 FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING P MAIN.
 FT DOMAIN 164 914
 FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
 FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
 FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
 FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
 FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
 FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
 FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
 FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
 FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
 FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
 FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
 FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).
 FT DOMAIN 3726 3730 COILED COIL (POTENTIAL).
 FT DOMAIN 203 292 GLN-RICH.
 FT DOMAIN 321 1010 GLU-RICH.
 FT DOMAIN 1846 2772 GLU-RICH.
 FT VARSPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT VARSPLIC 1637 1642 MISSING (IN ISOFORM 4).
 FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 4).
 FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 4).
 FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 3).
 FT VARSPLIC 2183 2907 MISSING (IN ISOFORM 3).
 FT VARSPLIC 2895 2907 MISSING (IN ISOFORM 3).
 FT VARSPLIC 2895 2948 MISSING (IN ISOFORM 5).

FT VARSPLIC 3901 3911 SITOPHAGRR -> ALBUTTSWHSNANI AIT HPEILSH
 FT VARIANT 1347 1347 SLG (IN ISOFORM 6).
 FT
 FT CONFLICT 76 76 /FTID-VAR 010926.
 FT CONFLICT 475 475 E -> Q (IN REF. 3).
 FT CONFLICT 554 554 M -> I (IN REF. 3).
 FT CONFLICT 638 638 E -> G (IN REF. 3).
 FT CONFLICT 663 663 R -> S (IN REF. 3).
 FT CONFLICT 913 913 N -> S (IN REF. 3).
 FT CONFLICT 956 956 H -> N (IN REF. 3).
 FT CONFLICT 980 982 K -> N (IN REF. 3).
 FT CONFLICT 997 997 OKH -> PKP (IN REF. 1 AND 2).
 FT CONFLICT 1001 1001 Q -> P (IN REF. 1 AND 2).
 FT CONFLICT 1020 1020 N -> D (IN REF. 3).
 FT CONFLICT 1028 1028 V -> E (IN REF. 3).
 FT CONFLICT 1626 1626 R -> P (IN REF. 1 AND 2).
 FT CONFLICT 1703 1703 N -> T (IN REF. 3).
 FT CONFLICT 1707 1707 V -> G (IN REF. 3).
 FT CONFLICT 1802 1803 MISSING (IN REF. 5).
 FT CONFLICT 1843 1843 A -> P (IN REF. 3).
 FT CONFLICT 1956 1956 I -> V (IN REF. 3).
 FT CONFLICT 2027 2027 V -> D (IN REF. 5).
 FT CONFLICT 2157 2158 EI -> HE (IN REF. 7).
 FT CONFLICT 2514 2514 E -> V (IN REF. 3).
 FT CONFLICT 2851 2851 L -> R (IN REF. 3).
 FT CONFLICT 2957 2957 I -> N (IN REF. 8).
 FT CONFLICT 2983 2983 E -> D (IN REF. 3).
 FT CONFLICT 3087 3087 P -> S (IN REF. 3).
 FT CONFLICT 3218 3218 O -> H (IN REF. 3).
 FT CONFLICT 3307 3309 O -> H (IN REF. 3).
 FT CONFLICT 3751 3751 ESE -> OSO (IN REF. 3).
 FT CONFLICT 3751 3751 P -> A (IN REF. 3).
 FT CONFLICT 3833 3833 T -> S (IN REF. 3).
 SQ SEQUENCE 3911 AA; 453664 MW; 3FB1CB1C819B47AA CRC64;
 Query Match 3 6%; Score 191; DB 1; Length 3911.
 Best Local Similarity 19.9%; Pred No. 0.12.
 Matches 222; Conservative 172; Mismatches 360; Indels 4; Gaps 54.
 Oy 41 EYKEQIORT-VATSGNMAVELL-STLRGWHLQMTREFEALRTGSI-AAATW 95
 1832 ELSQRLVSGFAGTEIDPENNELMINSRLQAAV-EKLEALSETSN-EHA 1883
 Oy 96 NEPLTDLPSRSPENADEFLQNLLOPTLVKLVROYIDKMBE-ELLITLNNRIA 153
 1884 KYVOTELMRESEFOKOEATEST-KCELELRERLHR-SKARV 1924
 Oy 154 AANNNGNSGV-BELKRIYKEMWESAFLNLRQ-TGNNELVQELISDS 203
 1925 LAVELSKACVLDGVADEKTLFEKOIEK-TDITLRLQELLCASNL-OFIE ARQ 1979
 Oy 204 ESNAELEIENS-QVQKQVEBOLLST-VOPNLKRWGME 241
 1980 QIQEHELLSRQKAKKAGVQDGLQETELKKKELEVQCAKAVYDQKQVKALE 2039
 Oy 242 NNSSESPADSSVSESPISLAEGSVCLDESIGNNSGMSGTGSDENAVARAS 301
 2040 IVEEEDVSFELEQEKNTLM-DLRQOQALEKOLERKFLDQV 2086
 Oy 302 PEPELQRLRYOMEVQPALEGKNIICLPTGSGKTRVAVYIAKDLHKKKASHYXVIV 361
 2087 DREHERDYFOOELQLEOOLVVPRPQISQHTREVEQALNNHLEKTKDKST 2140
 Oy 362 LVNKVLLV-EQLFRKEFQPLKKYRVVIGLSQDTQIKI-SPREVV 404
 2141 LLSKEDLOQDIERNEIEIKLEFR-VRELBALLVSAOTPKVYEDRKHFAGVIAKPEIS 2198
 Oy 405 KSCDIIISTAQT-LENSLNLNENGDACVQSDSLIIDCHHTNKNEAVYNNIMKHYI 462
 2199 LEVOQDAERDAIDRKEKTEITNEE-QEQR-ELEENKNEE-YOUIMOULEI 2247
 Oy 463 MOKLKNR-LKKEKPYIPOLPILGILASGVCATQAKA-EHHIILKANI 515

```

Db 2248 OKKESTTRLOELBOENKLFKDMKELGL-----AIKESDAMSTODQVL---FGKRA 2296
QY 516 FTTRTKKENLDOLKNOIOECKKFAADATREDPEKELL-ELMTIIOYVOMSPSDRG 574
Db 2297 QIIIOKEVEIDOLNEQYTKLOOQKITTDKVIIEKNEELRLEIOIE-CLMS----- 2348
QY 575 TQYEDMAIOEKKAAKKKGRVCAEHKRYNEALQINDTRMDATVILLETFYNEEK 634
Db 2349 -----DOECVNR-NREE-----ETEOLEVEI-----EKL 2371
QY 635 DKKFAVIEDSDGDEYCDGDEDEDLKKPLKLDETDFTLTFEENKMKRLAENP 694
Db 2372 QOELANIGOKTSMNANSL-----SEPADSLKHQDVIYAEL-----ALEQV 2414
QY 695 EYENKULKLRNTIMEQYRTRESANGIIFTKROSAYALISQMITENEKAEVGAHHL 754
Db 2415 ETANEEMTEKKNVLEKTFNEMNOLTOELFSLKRESEVEKIOSIPEN---SVNAVADHL 2470
QY 755 IGAGHSEEPK-----MTONEQKEVISK-FRT-----IN-----LITVAEEG 795
Db 2471 -----SKOPLELVLTEDALKSLENOYFKSPENGKSTINLETRLQLESTVSAKD 2524
QY 796 LDIKEC-----NVIYRGVLVTEIEM-----816
Db 2525 LELTQCKQIKDMQESQFETEMQKKIVNLOKIVEKVAALVSQIQLVNOEYAKFCQ 2584
QY 817 -----VOARGARADE-----STYLVVAHSGVTEHETVND 848
Db 2585 DNOJTISEPERTIOMLNQRLDELDGSDLSALTLRILESESQVEHMT-SLIEKEQV-E 2642
QY 849 FREKMYKKAICHYQNMKPEYAKHILELOMSIMEKKMKTKRNIAKRYKNNPGLITFLCK 908
Db 2643 IAEKNVL-----EKEKKLEL-QKLEGENEKORE-KEKKSPP-----2678
QY 909 NCSVLACSGEDIHVI-----EKMHVMTTPFKELIYVR-ENKALOKKCAVQINGELICK 963
Db 2679 -----QDVEYKLTTELPHSNESGPFNELEALRAVSATKALAYAKKAE-----2725
QY 964 CGOAMGTMYHKGIDLPCLKIRNFVYFKNNSTKKQ 999
Db 2726 -----KLOELLVKEKTNMTSLQ 2742

RESULT 15
CENT_HUMAN STANDARD: PRT; 3210 AA.
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CNP-F, kinetochore protein (Centromere protein F) (Mitsin) (AH
DE antigen).
GN CNPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; H. mo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RA MEDLINE-95348175; PubMed-7542657;
RX Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CNP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-95379848; PubMed-7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-T., Shan B.,
RA Jones D., Yang-Feng T.-L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic phase progression.";

```

```

RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE-95336446; PubMed-7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE-95370296; PubMed-7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE-98437347; PubMed-9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CNP-E reveals
RT interactions with the kinetochore proteins CNP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC - FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CNP-E AND BUBR1.
CC - SUBUNIT: HOMO- OR HETERODIMER.
CC - SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS);
CC REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF
CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
CC - DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
CC - PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U19769; AAA82889.1; -
DR EMBL: U30872; AAA82935.1; -
DR EMBL: U25725; AAA68889.1; -
DR Genew: HGNC:1857; CNPF.
DR MIM: 600236; -
DR InterPro: IPR001230; Prenyl_site.
KW Chromosomal protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1338
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT DOMAIN 3015 3032
FT CONFLICT 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561
SQ SEQUENCE 3210 AA; 367589 MW; 11DB3324960E4334 CRC64;

Query Match 3.6%; Score 189.5; DB 1; Length 3210;
Best Local Similarity 19.1%; Pred. No. 0.11;

```

Matches 218: Conservative 187: Mismatches 473: Indels 265: Gaps 46:

```

OY 21 RYKMYIQVEBVL-----DYLTFLPAEYKEQIQRTVATSG-----NMQAVELLIS 64
Db 1917 KIEACIELEKIEYELKENSNDSEKLEYFSCDHQELQVETSEGINSDI EHMADKSSRE 1976
OY 65 TLEKGYVHL--GWTREFEVLEARTGSPLAARYNNPELTJLPSPSPFNABH EYLOLINILQ 122
Db 1977 DIGDNVAKVNDKWERFLD-----VENELSRIRSEKASTIEH-IALYLPADLE 2022
OY 123 PTLVDEL-LVRDYLDK-----GMEELLTTI-EDRNRILAAENNGNE--SVRELLKRIYQ 173
Db 2023 VYGTETKIELEKONENKQXIVYCLEEBELSVYTSERNQLRGELDTMSKTTVILQLSKMMG 2082
OY 174 K---ENMFSAPLNLKQT-----GNELVQELTSGDSNSNAELENLSQ VDGPOVEHQ 223
Db 2083 KIOELESHOSSECHICIQVAAEAYEKEKTELLQTLIS-SDVSELKDKTHLQFKQSLKEDSQ 2141
OY 224 LLST-----VOPNLKEVWGMENNS-----SESSPADSSVVSERSDLSAEGS----- 266
Db 2142 ALSLTGCELENOIAOLNKKERELLYKESLQARLSSESDYFKLNVSKALEALVEKGEFAL 2201
OY 267 -VSCLDSELSGHSNMGSDSGTMSDDE-----ENVARASPEPELQI RPYOMEVAQP 318
Db 2202 RLSSSTQEV-HQLRIGIEKLVRVIEADEKKQHLIAEKLKREKNDLSLKI KVENLRFELQ 2260
OY 319 ALEGNIIICLPTGSGKTRAVAYIAK-DHLDKKKKASEP-KYIVLVNXYLVQELFKREP 377
Db 2261 MSEENQELVILDAENSKAEVETLQTQIEEMARSLKVFELVLTREKENITFKQIOEKQG 2320
OY 378 QPFLKMYRIVIG-----LSGDTQOLKISFPEVXSCDIITSTADLENSULNEN----- 426
Db 2321 Q--LSELDKLLSFKSLLEKEBOAEIQKEESK-----IAYEMLCNQLPLNEAVAAALG 2372
OY 427 GEDAGYOLDSFLLIIDECHTNTKEFAVYNNIMHLYLMOKLKNRKLKKNKPIVLPPI-- 484
Db 2373 GQGEIMKATEGSLDPRIEEHOLRNSI-----EKLRARLEADEKKQCLVLOQLKE 2422
OY 485 -----LGLTASPGVGATKQAKABEHILKT-----CANLD 514
Db 2423 SEHHADLLKGRVENLERELEIARTNOEHAALEAENSKGHEVETLAKIEGATQSLRGLEHD 2482
OY 515 APTIKTVKENLDQKNOIQEPCKKFAIADATREDPRKELLETNRIQTYQOMSPMSING 574
Db 2483 VVTRISEKEDLTNELOKEQERISELEIINSSPENILOEK -----EQRKYOMKEKSTPA 2535
OY 575 TOPPEQMAIOMEKKAKKNSRERVCA--EHLRYNEALJIN--DTIRMLDA-----YTHL 626
Db 2536 MEMLOTOLKELNRVALLNDQFACAKAQNLSOYEGLTLEKAOLOLOSI DPAKKNNYIVL 2595
OY 627 ETFFYN-----BEKDKKFAVIEDDSDEGDDHFGDCQDEDEDDI KKPCLKIEMTD 673
Db 2596 OSSVNGLIQVEDGKOKLEKKDEISRLKNOI-----QDOEOI VSKLSQVEGE 2643
OY 674 RLMLTLFFENNNKMLKLAENPEY-----ENELTKLKNIMEQYTRTHSARGI IFTK 726
Db 2644 H---QLMKQONLELRLTVLELQKIQVLOSKNASLQDTLEVLOSSTYKNLENE--LELTRK 2697
OY 727 TROSAAVLSQWITENEKFAVGVKAHLIGAGHSEFFKPMTONFOKEVYISKPRTG INLL 786
Db 2698 MDKMSFV-----EKVNMKTAKETELOREHMEHMAOKTAELOHLSGCKNRLAGHLOLL 2749
OY 787 -----IATVAEBGLDIKECNIVIRYGLVTNEI AMYQAKARADESTIV 831
Db 2750 LEEIKSSQOLKELTLENSSELKSLDCMHKDOYEKGEKVREEIAVEYOLRLHEAEKKHQAL 2809
OY 832 LVAHSSGVIEHETVNDPFEKMMYKAIHCVQNNKPE-----FYAKHILELOMSIME 883
Db 2810 LLDTNQYVEVLOT--YREKLITSKE-ECLSSQKLEIDILKSSKEELNNSIKATQO-ILE 2864
OY 884 KKKMTK-----RNIAKHYKNNPSLITFLCKNSVLAQSGEIHVIEK----- 925
Db 2865 ELKKTMDNLKYNOLKKEENRQAGKMLILSKOLE---EKEKELOKHLISOLOAAQEK 2921

```

```

OY 926 -----MHVNMTPERKELYIVREKKA-----LQKKCADYQINGEITK*2.AM:IM 971
Db 2922 QKTGYMDTKVDELTTETIEIKELKETLEEKTKADEYLDYGS-LIISHKELPKAK:MLH IQ 2980
OY 972 MVH 974
Db 2981 VAH 2983

```

Search completed: May 8, 2003, 15:53:32
Job time : 51 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:50:13 : Search time 52 Seconds

(without alignments)
1894.957 Million cell updates/sec

Title: US-09-515-363c-2

Perfect score: 5311

Sequence: 1 MSGXSTDENFRYLSCFRA.....LPITPNLDSECLFSDDD 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 687 | 12.9 | 1037 | 2 T32534 | hypothetical prote |
| 2 | 654 | 12.3 | 811 | 2 T30968 | helicase homolog C |
| 3 | 642.5 | 12.1 | 956 | 2 AB8708 | protein COL10.1 |
| 4 | 516 | 9.7 | 398 | 2 T46312 | hypothetical prote |
| 5 | 499 | 9.4 | 1134 | 2 T20332 | hypothetical prote |
| 6 | 453 | 8.5 | 752 | 2 D75219 | ATP-dependent RNA |
| 7 | 441.5 | 8.3 | 650 | 2 D71203 | probable ATP-depen |
| 8 | 417 | 7.9 | 741 | 1 A69432 | ATP-dependent RNA |
| 9 | 402.5 | 7.6 | 1374 | 2 S62524 | probable RNA helic |
| 10 | 392.5 | 7.4 | 784 | 2 H64487 | ATP-dependent RNA |
| 11 | 350 | 6.6 | 1822 | 2 S44849 | ATP-dependent RNA |
| 12 | 342 | 6.4 | 738 | 1 E69055 | K12H.4 protein - |
| 13 | 313.5 | 5.9 | 784 | 2 D84386 | ATP-dependent RNA |
| 14 | 303.5 | 5.7 | 993 | 2 S48436 | probable RNA helic |
| 15 | 284 | 5.3 | 502 | 2 T31323 | ATP-dependent RNA |
| 16 | 278.5 | 5.2 | 839 | 2 T39190 | probable ATP-depen |
| 17 | 223 | 4.2 | 557 | 2 G69958 | SNF2 helicase homo |
| 18 | 218 | 4.1 | 1979 | 2 C71622 | hypothetical prote |
| 19 | 217.5 | 4.1 | 2663 | 1 S28261 | transcriptome prote |
| 20 | 214.5 | 4.0 | 1790 | 2 S67593 | transport protein |
| 21 | 209.5 | 3.9 | 449 | 2 C71156 | probable helicase |
| 22 | 205 | 3.9 | 453 | 2 B75006 | RNA repair protein |
| 23 | 205 | 3.9 | 967 | 2 D72308 | conserved hypotet |
| 24 | 204.5 | 3.9 | 2163 | 2 S50675 | pre-mRNA splicing |
| 25 | 204 | 3.8 | 1939 | 2 T18372 | repeat organellar |
| 26 | 201.5 | 3.8 | 1063 | 2 T38420 | probable DNA helic |
| 27 | 198.5 | 3.7 | 2253 | 2 T30336 | nuclear/mitotic ap |
| 28 | 197.5 | 3.7 | 1163 | 2 G97236 | ATPase involved in |
| 29 | 195 | 3.7 | 1531 | 2 T48946 | hypothetical prote |

| | | | | | |
|----|-------|-----|------|----------|---------------------|
| 30 | 195 | 3.7 | 3259 | 1 A56539 | glutathione - human |
| 31 | 194.5 | 3.7 | 986 | 2 E90596 | residuos-modifi |
| 32 | 194.5 | 3.7 | 1256 | 2 T26101 | hypothetical prote |
| 33 | 194 | 3.7 | 1039 | 2 E27234 | hypothetical prote |
| 34 | 193.5 | 3.6 | 663 | 2 G89848 | exonuclease ABC su |
| 35 | 193.5 | 3.6 | 2017 | 1 A36014 | myosin heavy chain |
| 36 | 193.5 | 3.6 | 2057 | 2 S61477 | myosin II heavy ch |
| 37 | 193.5 | 3.6 | 3225 | 2 I52300 | glutathione - human |
| 38 | 193 | 3.6 | 715 | 2 B75135 | DNA helicase relat |
| 39 | 193 | 3.6 | 986 | 2 H90565 | restriction modifi |
| 40 | 192.5 | 3.6 | 718 | 2 E71100 | hypothetical prote |
| 41 | 192.5 | 3.6 | 1300 | 2 I53799 | Gcl protein - huma |
| 42 | 192 | 3.6 | 986 | 2 S49394 | HsdR protein - My |
| 43 | 192 | 3.6 | 1356 | 2 S37263 | kinectin-1 - human |
| 44 | 191.5 | 3.6 | 943 | 2 F65543 | ATP-dependent RNA |
| 45 | 191 | 3.6 | 1922 | 2 T00637 | hypothetical prote |

ALIGNMENTS

| | | |
|---|--------|--|
| RESULT 1 | T32534 | hypothetical protein F15B10.2 - Caenorhabditis elegans |
| C:Species: Caenorhabditis elegans | | |
| C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999 | | |
| C:Accession: T32534 | | |
| R:Wohlmann, P.; Murray, J. | | |
| submitted to the EMBL Data Library, December 1997 | | |
| A:Description: The sequence of C. elegans cosmid F15B10. | | |
| A:Reference number: 221186 | | |
| A:Accession: T32534 | | |
| A:Status: preliminary; translated from GB/EMBL/DBJ | | |
| A:Molecule type: DNA | | |
| A:Residues: 1-1037 <WOH> | | |
| A:Cross-references: EMBL:AF036696; P1DN:AA88350.1; GSPDB:GN00022; CESP:F15B10.2 | | |
| A:Experimental source: strain Bristol N2; clone F15B10 | | |
| C:Genetics: | | |
| A:Gene: CESP:F15B10.2 | | |
| A:Map position: 4 | | |
| A:Introns: 30/3; 89/1; 111/2; 218/1; 258/2; 288/1; 343/3; 370/4; 400/2; 443/1; 520/1; | | |
| Query Match | 12.9% | Score 687; DB 2; Length 1037; |
| Best Local Similarity | 25.0% | Pred No. 1.4e-27; |
| Matches 251; Conservative 169; Mismatches 408; Indels 174; Gaps 35; | | |
| OY | 101 | DLRPSFENAHDE--YLDLNLLOPTLVLDKLVLDKCMEE----- 142 |
| DB | 63 | DLKNSILSNADDERLYKDIIMTYLQ--TYLPKCTVHKLLN--CSNREYKLSDFHYLDHFEGF 120 |
| OY | 143 | LITIEDRNRIAAENNGNSGVALLKRIYOK---ENMTSAPLN--VLFQIG--NNELVQ 195 |
| DB | 121 | LRFTEKRYVLAIDSYPOYIDAVAVLRKEIEREEDNODSFKKLITLTVIGLQAVY 180 |
| OY | 196 | EL--TSDSCSNAEILENISQVNGPQVEPOLISTVOPNLEKVMGMEN-----N 243 |
| DB | 181 | DIWYTTSEKSSNNLDVE-----AKOFIA-----KVLAKMDGKREYQIIN 221 |
| OY | 244 | SSESPADSSVSESPTSIAEGSVSCLDESLGNS----NMGSDCTMGSDNI E--ENV 296 |
| DB | 222 | ASRQRLNGRIYICPVHESATEMMVYLGTAALNTRFMINIRVDNI VOENSIHRAVIESV 281 |
| OY | 297 | AAASPEPELOLPYOMEVAOAPLEKNIIICLPTSSGTRAVYIAKHILKKKKASEP 356 |
| DB | 282 | RQRIHQRODCLFNVOEELQVALQGNIVTAPTSQGTVAANLIKHFYRSSEGR 341 |
| OY | 357 | GKVIYIV--NKVLVLEOLFKEFQPLFKMYRVIGLSGDTQKISPEVYKSLIITSTV 415 |
| DB | 342 | FKALFPTPMSMLINQ--AAISISYLDHYHQQIIGSD--NVPRTNVLSQKLLIVATQ 397 |
| OY | 416 | ILENSLLENLNGEDAGV-----LSDFSLLIIDECHHTKKAQVYNNIMKHYI HOKLKNR 470 |
| DB | 398 | MIYNLCNHNHNSLDDSRILDDQFLSTFTIIFDECHNTVKNPSYNNIMREY--HYLKNNG 455 |


```

Db 1 LESTYRPEKGEFSELLPLGKIDELKTHS-ENSGEFSKQLYDQLKSN:SIADP-----E 75
Oy 129 LYRVDVL-----DKMEBELITEDRRIRIAAENNGNSCVRELLKRIYOKENWFSAF 181
Db 76 RLYXDIAIYLQSNFPCILIDYLVNYSGRN-VLSEFRYILDLBEGFLRJIIPK-----VV 129
Oy 182 LNVLRGTGN-NELVOELTSGDSSESNAELENJLSDYDQPCVEGELSTTVJLNLKEWJWJ 240
Db 130 LNVLDSTPOYSEAVKLL-KDEIDRIEEDNN-----DPEFIKLLIRTVILGDAVYVIT 182
Oy 241 ENNSSESSPADSSVYSESDTSLA-EGSVSCDLESGLHNSMGSDSGTMSDSEZNAAR 299
Db 183 MYTIWEKAQMTSIIINIRVDSIAPENSASRL-----VIESVGR 221
Oy 300 ASPEPELOLRPYQMEVAOPALECKNIITCLPTGSGCTRAVAYIAADHLDKKKKASPEKV 359
Db 222 IHIORQCLKNRYOBELOQVALQCKNTIYVAPGSGKTIVAIANIKEHRESR---SSEG-- 276
Oy 360 IVLVNKLVLVEOLFKEFEOPFLKMYRVLGSDOTJLKISPEYVKSDIILISTAOILEN 419
Db 277 -----KRRKTQTOIGSD--NPTFNIVYQSDULVATPQMLVYN 311
Oy 420 SLNLENGEDAGVO-----LDSFLIITIDECHTNKEAAYNNIMRHYLMOKLKNNEKL 472
Db 312 -LGN-EHRPDLDDEYPEPEOFLSTFIIFPEDENHTVNSKSPYSNMYREY-HYLMKNMGM 367
Oy 473 KENKPVILPOLLGLTASPCVGATQQAABEHLIKGCLNDAFTIKVAKINLDQKJL 532
Db 368 PEGH---SFOIITGLTASLGTGKKMKCMQYRSTIAGLCNNMYKELSTYKONLEELHYN 424
Oy 533 QEPCKFAIADATREDP-----FKELLETIRIQYQMSPMWDFQJLP-----Y 578
Db 425 PFTVDQYFCEFSRSDGPIEMFTRKLQMOQEVEDLIRTLKNEP YVKEHPEPTOKENRY 484
Oy 579 EOMALQMEKKAARKGNKEVCAEHLRKYNEALQINDIRIMDATYHLTYNEKEKKEFF 638
Db 485 ENMILSNORNCVSLAGSNKLLIIEVL-----DVLKQYMSQMHITWYFRKTRF 534
Oy 639 AV--IEDDSEGGDEYCDDEDEDLKRPLKD-----ETDRFLMILFEENNRKLK 688
Db 535 GKFRIFOFOCFYVALSYNINPEVALMKILKEKLEPERIRNFTON--NLIIMDN--CHR 590
Oy 689 RLAEINPEYENKLLKLRNTIMQYTBTEBAROIIFTRKQSAVALSONIENKFAVYG 748
Db 591 ELVIGSAENPMIARIVQFLLDQNEOTSD-FRAIIFVRKKADEFLAVYL--NDRLHELG 647
Oy 749 VKAHLHLG--AGHSSEFKPMTONEKEVYSKRTKIMLILATVIAEGLIDIKCNVI 805
Db 648 IKSOMSGOKKSTASSADIASAKOMKEMIKMADEBNQILVSTVAEGLDIDPECSLVI 707
Oy 806 RYGLATNEIAMVQARGARADESTYVLVAHSGSVLEH--ETVNDREKMYKAIHCYON 863
Db 708 KYNAVTEDETAHVORGRARARANSKCVLITS--ILHVOESNNLAKENLMTETISLON 764
Oy 864 MKPE-----FYAKITELQMSIMEKMKTKENIAKHKNPSLITTECKNGCSVLASG 917
Db 765 SPGEFRQCVDESKWYPRIOREDTDAQRIRKQI-----NNRIYIKYCMKGDIVLCTN 819
Oy 918 EDIHVIEKMHVNTPEF---KELYIVRENKALOKKADYOJNEIIC--KCGQAMGT 970
Db 820 KDIR-SKNTQIYVNCPEGMSLVYRIRIPLERQASRN---FNSTGSIECIGRSGSLQ 874
Oy 971 MMVHKGLDPLKIRNFNVVYVFNKNSNRKKOYKKKVEL-----PIFPNLI 1014
Db 875 LIDVNVNLPCLAKVSTILLIESTNEKILVQKKNKILDLHPFTPTLQKRI 924

```

RESULT 4
T46312
hypothetical protein DKFZp341J111.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence-revision 04-Feb-2000 #text-change 04-Feb-2000

C:Accession: T46312
R:Diesterheft A.; Lauber, J.; Mewes, H.W.; Well, B.; Wiemann S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23036
A:Accession: T46312
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-398 <AA>
A:Cross-references: EMBL:AL137608
A:Experimental source: adult testis; clone DKFZP43A1111
C:Genetics:
A:Note: DKFZP43A1111.1

```

Query Match Similarity      9.7%: Score 516; DB 2; Length 348;
Best Local Similarity      33.3%: Pred. No. 2,2e-19;
Matches 141; Conservative  68; Mismatches 157; Indels  5; Gaps  14;

QY  594 NKRKRC-----AEHLKRYNEALQINDTIRMIDAYTHLETFYNEKKKRAVITGHSDE 647
      : : : : : | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db   2 DEESIRICALFLYTSHLKRYNDALLISEHARKDALDYLKOFFSNVRAAG----- 52

QY  648 GSDDEYCGDDEEDDLKKPKLDEDRPLMTLEFENNMYKRLAENPEYNEKTLNINI 707
      : : : : : | | | | | : : : : : : : : : : : : : : : : : : : : : :
Db   53 -----DELEODL-----TOR-----FEKKLELESVSOPSNEKKLDFI 90
      : : : : : | | | | | : : : : : : : : : : : : : : : : : : : : : :
QY  708 IMEYRTTEESARGIIFTROSAYALSOMITENER--FAEYGVKAHHLIGASHSEKRP 765
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db   91 LQEFYHLNPETIT-ILFVKTRALVDALKNMIGNPKLSPLKGI-----LTGRKTNUNG 145

QY  766 MTONQKVEISKFR-TGKINLLIATTVAEEGDIDIECNVITRYGLVTNIEIAVQAKHAR 824
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db   146 MLLPMAKCTLLDFKASKGDMNILLIATVADEGIDICQNLVILYEVGNIKKIKIIRGGR 205

QY  825 ADESTYVLVAHSGSVIEHEITVNDPREKKMYAIIACVQNMKDEEYAHKILLEOM--INKEK 884
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db   206 AGSGSCFLLT-SNAGVIEKEQINMYKEKKMNDISIRLQTDWDAVREKRIIHLIT-- 261

QY  885 KKKTRNIAKH--KNNPSLIFFLCKNSVLACSEDIHVIEKMHVNNITPREKELIVR 942
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db   262 FLRDSOEKRPYPDDENKLL--LDRKCALACYADRAVDIECHTYVLGDPAKESVSR 318

QY  943 ENKALOKKADYQINGELIC---KCGQAWGTMMVHGGLDPLCKIRNFVYVFRNNSKKQ 999
      : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db   319 PPRK-PROGSSPEKRAKIFCARONCSHDGIMHVKKTFEIPVLIKIESFVEIILATVLL 377

QY  1000 YKKW 1003
      : : : : :
Db   378 YSKW 381

```

```

RESULT 5
T20332
hypothetical protein D2005.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20332
R:Wilkinson, J.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19258
A:Accession: T20332
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 11134 <MIL>
A:Cross-references: EMBL:Z29752; PIDN:CA802082.1; GSPDB:GM00017; CSE:Z2005.5
A:Experimental source: clone D2005
C:Genetics:
A:Gene: CESP:D2005.5
A:Map position: 1
A:Introns: 33/2; 91/2; 116/3; 144/1; 319/1; 361/1; 421/3; 482/2; 507/1; 684/3; 757/1.;

Query Match          9 48:  Score 499;  DB 2;  Length 1144;
Best Local Similarity 23.58;  Pred. No. 7.3e-18;
Matches 193;  Conservative 134;  Mismatches 288;  Indels 20;  Gaps 29;

```


DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon

A:Reference number: A71000; MUID:98344137; PMID:957194

A:Accession: D71203

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-650 <KAW>

A:Cross-references: GB:AP000007; NID:93236134; PIDN:MAA1011; PID:91258336

A:Experimental source: strain O73

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1896

C:Superfamily: ATP-dependent RNA helicase eif-4A

C:Keywords: ATP; nucleotide binding; P-loop

F:133-40/Region: nucleotide-binding motif A (P-loop)

F:131-136/Region: nucleotide-binding motif B

F:135-138/Region: DEAD/H motif

Query Match

Best Local Similarity 25.9%; Pred. No. 36-15;

Matches 185; Conservative 122; Mismatches 261; Indels 147; Gaps 29;

8.3%; Score 441.5; DB 2; Length 650;

306 LQRPYQVMAOPALEGKNIITCLPTGSKTRVAVYIAKHLLKKKASHFGKAVLYLV-N 364

11 LQRPYQ-ELIYAKKERKNCILVPLTGLKTIAMMADYRLIK-----YCGKVLMLAPT 64

365 KYLVVE--QLFRKQPPFLKKWYRVIGSGDQKISPEVYKVS---CDIISTAGILEN 419

65 KPLVLDHAETFRKFFTLPPK-----IYALTGE-----ISPMRVAMARAVVIATPOTIEN 117

420 SLINLENGEDAGVQSLDLSIIIDECHTNKAEVYNNIM-HYLMQKLNKKKKENKPV 479

118 DLL-----VGRSLDVLIIIDFNAHRAVGNVAVIYIAEYIKO-AKN----- 159

480 PLPQILGLTASPVGATGAKAEHILKLANLAFITLYKENDOLKNOIOEPCKKF 539

160 --PHVIGLITASP-----STPEKIMEVLRMLGIEHI-YRSENSPDVKKVYQV--TRF 207

540 AINDATREDPEKE--KLIEIMRIOTYCOMSPKSDFTQIYEDMALOMEKKAACKNR-- 595

208 EAWKVELPELYKEVRKLRDMR-----DSKPLAEAGL--LDSTADIPKKEILRADQI 261

596 -KERVCAEH-LRK-----YNEALOINDTIRIMDAY-----THLETYNEKDKKFAVID 643

262 NEEMANGNDLRLKLLFLHMAKLHAIIELETOGLSALKAVYIKRLEYEAK----- 312

644 DSDGDDDEYCDGDEDEDLKKPLKIDETDFLMTLFFENNMKLRLEANPEYENKLIK 703

313 -----ACSTRASKELFLDKRMKKKALLIIO-----AKELGIDHFKMEV 350

704 LRMTIMEQYTRTESARGITFTKTROSAYALSOMITENKFAVGVKAKHLIAGISSEF 763

351 LKGLIRFQRLK-RKNSKIIVFTYRTAKRV-----EKLMKQIKAKRIVGQASREND 403

764 KPTONQOKEVISKFRITGLINLIATTVAAEGDIKECNIVIKYGLVTN-TAVOAKR- 822

404 RGSQKQKILIDAFARSEPNVLAISVGEGLDVPEDVLIVTEVPYPS-RSVQKQKQI 463

464 GRCQKPRVILMAQTRDEAVYMSRQKIKIMETIRNVSQ-----VRHRLQSLSESV 518

823 ARADESTYLVLAHSGSVLEHETVNDFERKMYKATHCVANMKPEEVANLDELQMSIM 882

464 GRCQKPRVILMAQTRDEAVYMSRQKIKIMETIRNVSQ-----VRHRLQSLSESV 518

883 EKMKAKTRINAKHYKNNPSLITFLCKNGSVLACSGEDIVTEKMHVNNIPEKEL----- 938

519 KREVNEE---IKKEKKEGTGKIVKIDSRILRS-----EVKKIKLTLGIGIEVRLILEVD 569

939 YIIRKNAALOKKCADYOING-----ELICKGQAWKMMVH 974

570 YIYSEDAIERSANDFOSIIDGRFLDQYKRLKEAVPRVITILEGOLVIRNVH 624

RESULT 8

A69432

ATP-dependent RNA helicase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-May-2001

C:Accession: A69432

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.F.; Ketchum, K.A.; Dod

: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.C.; Gill, S.; Kirschner, E

: Glöck, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

: Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,

: Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic archaeon

: Archaeoglobus fulgidus

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: A69432

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 <KLE>

A:Cross-references: GB:AE001002; GB:AE000782; NID:92689325; PIDN:MA89786.1; PID:9264

C:Superfamily: ATP-dependent RNA helicase eif-4A

C:Keywords: ATP; nucleotide binding; P-loop

F:132-43/Region: nucleotide-binding motif A (P-loop)

F:132-137/Region: nucleotide-binding motif B

F:136-139/Region: DEAD/H motif

Query Match

Best Local Similarity 23.8%; Pred. No. 6-5e-14;

Matches 169; Conservative 114; Mismatches 230; Indels 136; Gaps 30;

7.9%; Score 417; DB 1; Length 741;

303 EPELDQRYQVMAOPALEGKNIITCLPTGSKTRVAVYIAKHLLKKKASHFGKAVLYLV 362

11 EPELDQRYQ-ELIYAKKERKNCILVPLTGLKTIAMMADYRLIK-----YCGKVLMLAPT 64

363 VAKVLVLEQPFKQPPFLKKWYRV--IGSGDQKISPEVYKVS---CDIISTAGILEN 419

65 APTKPLVEDQAR-----FLKRYLAKVEIYLSGEVPEP-KKELMEKARIYSTQVVEN 118

420 SLINLENGEDAG-VQLSDLSIIIDECHTNKAEVYNNIMRYHLMQKLNKKKKENKPV 478

119 DLL-----AGRISLEVDILVPEAHRAVGNVAVIYIAEYIKRATK- 160

479 IDPQILGLTASPVGATGAKAEHILKLANL-DAFTTKT----- 520

161 -----PLIATASP-----SDPERIMEVIONGLIEIVRTQWSDVAVYKRRLE 209

521 -----VKNLIDLKNOIOPCKKPAIADTR-----EDPFKEKLEIMTRIGVYVNSP 569

210 WIKVDIPPEKAKVEKRLK-CIKIRKRLREIMIEVPENSSKRDIALQELADIAQASQ 268

570 MSDPQTPYQVMAOPALEGKNIITCLPTGSKTRVAVYIAKHLLKKKASHFGKAVLYLV 627

269 SEEL-----FEALSIILAE-----IMKLOHAEVLEITQGVKAVKILR- 305

628 TTYNEKKKKFAVIEDSDGDDDEYCDGDEDEDLKKPLKIDETDFLMTLFFENNMKLR 687

306 -----KLIVREATSKG----- 320

688 KRLAEND-----EYENKLTKLRTNIMEQYTRTESARGITFTKIQSVALS 735

321 KSTVGPPIKKAVALSKCKVEHPLEKLEIKLQEFENPNPS-RVYFTNHCENML- 378

736 QMTTENEFAEVGVKAKHLIAGHSSEFPQNOEKEVYSFRTGKINLIATTVAAEG 795

379 --VNELSPLEPVA-----KRVGQASRNDKGMQKQKQIETIDKFRGVYKAVIATN-IFG 432

796 LDIKECNIVIRGLVTNELIANYOARG-ARADESTYLVLAHSGSVLEHETVNDFERKMY 854

433 LDIPSTDLVFEAVYSEIRAIQKRGTRGREGRIVLVLTGK-----HDAV 481

855 YKAIHGVQMMKPEEVAN-KILELOMSIMEKMKTRKINAKHYKNNPSLITFLCKNGSVL 913

482 Y-----YSSMKKERKRYDILLET--KRILDK-----QRSIGDVLDEETGIVYVNSPEK 531

914 ACSGEDIHVIEKMHVNNIPEKEL-----YIIRKNAALOKKCADYOINGEI 960

Db 532 S-----EVYKHLREIGAKIEIRNLEAVDYVSDRAVERKTYHDELNLSI 576

RESULT 9

Probable RNA helicase/ribonuclease SPAC8A4.08c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000

C:Accession: T39130; T41432; T41192; S62524

R:Lye, G.: Church, C.M.; Batteil, B.G.; Rajandream, M.A.; Walsh, S.V.

Submitted to the EMBL Data Library, November 1995

A:Reference number: 221830

A:Accession: T39130

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1374 <LYE>

A:Cross-references: EMBL:266569; NID:q1052533; PIDN:CA91518.1; PID:q1052539; SPDB:SPAC8A4.08c

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Batteil, B.G.

Submitted to the EMBL Data Library, October 1998

A:Reference number: 221993

A:Accession: T41432

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1374 <SEE>

A:Cross-references: EMBL:AL032824; PIDN:CAB37423.1; GSPDB:GN0068; SPDB:SPCC584.10c

R:Lye, M.; Rajandream, M.A.; Batteil, B.G.; Seeger, K.; Harris, D.

Submitted to the EMBL Data Library, April 1999

A:Reference number: 221976

A:Accession: T41192

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 722-1374 <LYN>

A:Cross-references: EMBL:AL049662; PIDN:CAB41233.1; GSPDB:GN0068; SPDB:SPCC188.13c

C:Genetics: <LYE>

A:Gene: SPAC8A4.08c

A:Map position: 1

C:Genetics: <SEE>

A:Gene: SPCC584.10c

A:Map position: 3

F:32-39/Region: nucleotide-binding motif A (P-loop) #status atypical

F:141-146/Region: nucleotide-binding motif B

F:145-148/Region: DEAD/H motif #status atypical

Query Match 7.6%; Score 402.5; DB 2; Length 1374;

Best Local Similarity 25.6%; Pred. No. 8.8e-13;

Matches 175; Conservative 107; Mismatches 244; Indels 157; Gaps 32;

Db 304 PELQRPYQMEVAPALAEGRNIIICLPTSGSKTRVAAYIAKHDLK-----KKA 353

9 POL-LRKYQDDVYNNASK-QNTLLVMRTGAGKLLAVKLKOKLEQIILIGESNLEHKKI 66

354 SEGGKIVLVNKKLVQ--LFRKEPQPLKMYRVNIGLSDTQKISHPEVYKSDTII 410

67 S-----VFLNKKVPLVFOQAEYIRSQLPAVGMFVGEISLSENGEJC-----TINILKANYI 118

411 ISTAQLIENSLNLNGEDGAGVQLSDPSLIIDECGHTNKKEAYNNIMKHNYLMQKLNKK 470

119 VITADLF---YFLANG---FLSINDNLITFDECHAIAGNDAYARIMNIPY---HR 166

471 LKKE-NKVPILPOLIGLTASPGVGATQAKAEHILKLCANLDAFTIKTYENLDOLK 529

167 AKAVLSKKHHTLPRIQMTASP-----FTCK--KGNLYHHL 200

530 NOIQEPCKKFAIADATREDPFKEKLEIMTRIOTYQCMSPDSFQTPYHMAIQMKKA 589

201 YQ-----MEQLFDSKAHV---VSENLADYFCLLPRESVYMWYSNKLIV 238

590 AKGNKREKVCAGHL-----KRYEALQINDTIRM-----IDATYHLFFYNEPKDKK 637

239 VPPSDSITIKCELTQCKLISRAVKTALA--ETIDMGIMFGQVWLVIWPEVETKLRKK 296

638 FAVIEDSDGDDGDEYCDDEDEDLKKPLKIDETDRFLMTLFFENKKMKRLAEHPEVE 697

Db 297 KALGQLSD-----DEE-----LAIDRLKIFVED--WKNNKSDSNKPIVFD 337

698 N-----EKLTKRLMTIMEOYTRTEESARGIIFRTROSAAVLSOMITENKFAVWAKHH 753

338 SDYDVKRKLLELKATY-RKSDSVRYIPEKATATTSLEFM--KTLNEMIKHKS 393

754 LIGAGHS--SEKPMTONOKEVISKERTQINLLIATTVABEGDIKCONVIRYIVT 811

394 FIGHGSDGEGES-MTFRRQKDTLHKFKGYNVILATVABEGIDVDSNIVIMHNR 452

812 NELAWQARGRRADSTVILVAHSGSYIE---HEVYN-----DREKMYKATICY 861

453 TVTQVOSRGRKAAKSKFLIFLNTFEEILHRIIEEKNKLFALSENSNTHSIVCE 512

862 QNMK-PEEYAHILE-----LOMOSIMEKKMT-KRNIAHY-----KNINSLIFLC 907

513 EERRVTDLYVEGEGALLTGLYAVSLNYNCNLSRQVYTRYPTFAUQVSNWCF 572

908 -----KNCVSLACSGEDHIVIEK 925

573 EVELPCKCKVPAAGSPAKSIRK 595

RESULT 10

elf-4A family probable ATP-dependent RNA helicase homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-May-2001

C:Accession: H64487

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.; Sutton, G.G.; Blak, R.; Reich, C.I.; Overbeek, R.; Kinkness, E.F.; Weissbrock, K.G.; McInerick, J.M.; Gilek, I.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hult, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: H64487

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-784 <BD>

A:Cross-references: GB:067591; GB:L77117; NID:q1592136; PID:q1592139; TIGR:MJ1505; PI

C:Genetics:

A:Map position: REV1478605-14786251

A:Superfamily: ATP-dependent RNA helicase elf-4A

C:Keywords: ATP; nucleotide binding; P-loop

F:41-48/Region: nucleotide-binding motif A (P-loop)

F:139-144/Region: nucleotide-binding motif B

F:143-146/Region: DEAD motif

Query Match 7.4%; Score 392.5; DB 2; Length 784;

Best Local Similarity 22.9%; Pred. No. 1.3e-12;

Matches 169; Conservative 128; Mismatches 235; Indels 205; Gaps 28;

Db 306 LQLRPYQMEVAPALAEGRNIIICLPTSGSKTRVAAYIAKHDLKKKKSEKRVIVNK 365

19 LEARLYQQLIAAMALKKTKIV-ISTGDKTAIALVLAGLITRK-----DCKVILHLS 72

366 VLVLEQLFKKEPQPLKMYRVNIGLSDTQKISPEVYKSDIISTAQLIENSLNLE 425

73 RPLVEQHNRLKQVNLINDEKIALTGKIQK-KAELYKKGKFIATPQVYENDI---- 128

426 NGEAG-VQLSDPSLIIDECGHTNKKEAVYNNIMKHNYLMQKLNKKKKNPVPIHQL 484

129 ---AGRLNVDFTLLIDEAHHTTGDAH-----AFVAKKPKOK-----CHI 167

485 LGLTASPGVGATQAKAEHILKLCANLDAFTIKTYENLDOLK----- 529

168 LGLTASPG-----SDIKVMEICENIGIEHVEVREDEDEDKPYIAKVKLIPRIID 219

530 NOIQEPCKKFAIADATRE-----DPFEKLEIMTRIOTYQCMSPMDV 573

Db 220 PNEFKALK - LINEALKERLKLNDAGVINSIADVTTELIELNKKLFY----- 268
 Oy 574 GTOPYQOMAIQMEKKAAGKGNRERVCARLKYNEALINDITRMIDA-----YTHLE 627
 Db 269 -----DEEYKTELIVKCSALKLMAKKELESQKSVELANTN 306
 Oy 628 TFYNEEKDKFAVIEDSDSGDDEYCDDEDEDLKPLKLDETDRFLYLFENNKKM 687
 Db 307 KLSMOTKSAKSLVND-----KREAANLNLKSDVEHPKL- 342
 Oy 688 KRLAENPEYENKTLKRLNTMEQYTRTESARGIIFTRTSAYALSQYIENEFKFAV 747
 Db 343 -----KRVYDMVNNILEK-----NKDERIIIFQAYROIVETKVNITQN----- 381
 Oy 748 GYKAHLILGAGHSSEKPTQONOKEVISKFRGKINLIATVAEGFIKECNIVIRY 807
 Db 382 GIALRIFQCA-NKEGKSGOKQDAIEPRFK-EGSVLAVTSVSEGITPSVNTILRY 439
 Oy 808 GLVTNEIAMVOAGRARADE--STVYLVASHSGSVIENETVDFREKMMYKAIHCYQNN- 864
 Db 440 EAPVSEIRFIQRGRAMRGEGKVYVLLAKGTADNAVYSAL-YKERERKRL- KMG 495
 Oy 865 -----KPEYAHKILELQMSIMKMKTKRNIAKHK-----NNSLITFLKN 909
 Db 496 YLLNKLQKKFEKSEKEIEKEIEKEIEKESKTAVKEETKEEETKKTITLDFIKQ 555
 Oy 910 CSY-LACGED-----IHVIEKMHVNNTP-----EKKL-----YIVRE 943
 Db 556 IEKERSKEEDIKOEIKIPKKPIKTIYDVERKMAKLLHNYANIELKILEVDVYSD 615
 Oy 944 NKALOKKADYOINGEI 960
 Db 616 RYVERKTAEDFYNSII 632

RESULT 11
 S44849
 K12H4.8 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 30-Jun-2001
 C:Accession: S44849
 R:Favella, A.D.
 Submitted to the EMBL Data Library, May 1993
 A:Description: Sequence of the C. elegans cosmid K12H4.
 A:Reference number: S44851
 A:Accession: S44849
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1822 <FAV>
 A:Cross-references: EMBL:L14331; NID:g289702; PIRN:AAA2101...; PID:g289703
 C:Genetics:
 A:Insertions: 17/3; 72/1; 116/3; 170/1; 216/3; 340/3; 401/3; 413/3; 512/2; 634/3; 694/3; 74
 C:Keywords: ATP; nucleotide binding; P-loop
 F:33-40/Region: nucleotide-binding motif A (P-loop)
 F:141-146/Region: nucleotide-binding motif B
 F:145-148/Region: DEXH motif

Query Match 6.6%; Score 350; DB 2; Length 822;
 Best Local Similarity 23.5%; Pred. No. 6, 5e-10;
 Matches 175; Conservative 118; Mismatches 249; Indels 202; Gaps 37;

Oy 309 RPYOMEVAOPALEGKNIICLPTGSGKTRVAVYIAKDH-----IDKKKASEPKYIV 361
 Db 14 RDYQVELDKRATK-KNTIYVLGSGSGKTFYAVLLLEKYGQULPAPLDQCK-----RAPP 67
 Oy 362 LVNKKVLLVEOLFKEKQFPLKKWRYVGLSGDT-----OLKISFPEVYKSNQIITISAO 416
 Db 68 VVEKVNINVEQ---QAIHIEVHTSPKQGVHAGQTSGLMDSEKCEQKHHVAVITAGC 124
 Oy 417 LENSLLNLENGEDAGVOLSPSLIIDECHNT-NKEAVYNNIMRHVLMOKLKNRLAKEN 475
 Db 125 L-----LDLIRHAYLEKIEDMCVLFEDECHHALGSGHPRSIMVY-----KLLKRD 170

Oy 476 KEVILPQIIGTASPGVGATQAKAEHLIKLCANLDAFTIKTKVENLQIKNOLEP 535
 Db 171 KVV-----PRYLGITATL-IAKAVAPKELMBQKLLESANDS-VIETASD-LVS-SKAYGKP 224
 Oy 536 -----CKFAIA--DATREDFEKKLEIMTRIOTYQMSPMSPDQO----- 576
 Db 225 YEVVILCKFEELGCGIENFDVIEIPDETAVFVNTTEFHPDDLDRPRPKESLKITR 284
 Oy 577 -----PYEQALQMEKKAAGKGNRERVCARLKYNEALQIND-TIRMD-AYTHLE 627
 Db 285 AVFROLGPMAN-----RTQVMEKELGKTIKSOVLDPDKTLRFIMAKISMI 331
 Oy 628 TFYNEEKDKFAVIEDSDSGDDEYCDGDEDEDLKPLKLDETDRFLMIFENKKM 687
 Db 332 TI-----KRL-----EPMKKIKISLEALRPVVPQVIAFL 365
 Oy 688 KRLAENPEYENKTLKRLNTMEQYTRTESARGIIFTRTSAYAL-----SQWITE 740
 Db 366 ETF--NPEFOKERM-KLEKA-----EHLSAIIFVQRYAVYSLMLMHKISW-- 410
 Oy 741 NEKFAVGYKAHLLGAG-----HSSEFKPTQONOKEVISKFRGKINLIATVAEFL 796
 Db 411 EPRFK--FVNPDYVWAGSRLASSDSQGL-HKQTEVLRFRHNEINCLIA:SVLEFSV 467
 Oy 797 DIKECNIVIRYGLVTNEIAMVOAGRARADESTVYLVASHG-----SGVIE 842
 Db 468 DYKQCNLYKPRPLDMKRYVOSKGRARAGRYIYIEBKTAAYCSKLPDITRIVP 527
 Oy 843 HETVNDFERKMMYKAIHCYQNNKP--EEYAHKILELQMSIMKMKTKRNIAKHKNNP 900
 Db 528 HNOIIPFENGVTK--YCAELLPLINSPIKHA1-----VLKNPMNKTI----- 569
 Oy 901 SLITFLCKKCSYLACSGE-DIHVIEK-----MHVNNTP-----EKK 937
 Db 570 AQMAVALEACROLHLEGELDNLPRGRESIAKLEHIDEDEPYAPGIAANVSSKKRQ 629
 Oy 938 LYIVRENKAL-----OKKADY 954
 Db 630 LYDKRIARALNSESFEADKECFIY 653

RESULT 12
 E69055
 ATP-dependent RNA helicase, eIF-4A family - Methanobacterium thermoautotrophicum (str
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-May-2001
 C:Accession: E69055
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: E69055
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-738 <MTH>
 A:Cross-references: GB:A600903; GB:A600666; NID:g2622514; PIRN:AA85892.1; PID:g2622
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1415
 C:Superfamily: ATP-dependent RNA helicase eIF-4A
 C:Keywords: ATP; nucleotide binding; P-loop
 F:38-45/Region: nucleotide-binding motif A (P-loop)
 F:134-139/Region: nucleotide-binding motif B
 F:138-141/Region: DEXH motif

Query Match 6.4%; Score 342; DB 1; Length 748;
 Best Local Similarity 24.7%; Pred. No. 4, 6e-10;
 Matches 172; Conservative 106; Mismatches 234; Indels 184; Gaps 31;

Oy 302 PEPEIDLRPYOMEVAOPALEGKNIICLPTGSGKTRVAVYIAKDHIDKKKASEPKYIV 361

Db 12 PE-KIEARTYQOALLADYIRKNGSMIVAPTALAGTIVAVLAERLKKYMS-----KVLI 66
 QY 362 L-VKKVLLVQDLFRKEQOPFLKKRYIGSGTOLKISF-PEVVK-----SCDIIISTAQ 415
 Db 67 LSPSKPLAIOH--EESFREEM-----LATCSILGSIKPEKRRKRIKSOISATPO 116
 QY 416 ILNLSNLNLENGEDAG-VQSDFSLLIIDECHHTNKFAVYNNIMRHYLMOKLNRRKKE 474
 Db 117 TVESDIL-----AGRYDLRDVSLYFDECHRAVGSYVFLASNT-----IQMAR----- 161
 QY 475 NKPIVPLPOLGLTASGCGATKQAQAEHILKLCANL--DAFTIKT-----VKENDIQ 527
 Db 162 -----HPLILGLTASGAD-----EDKIKTVCENLPMNEVYVTECHIDPVPYLPK 207
 QY 528 LKNOIOEPCKKFAIADATREDPFREKLEIMTRIQTCCQSPMSDGTOTVEOMAIQMEK 567
 Db 208 IKLEIWKV-----VRRTPLEEDRELRLRVLKN-RKMLKNLGVII-----TISVCK 252
 QY 588 KAARKG-----NRKERYCAEHLKRYNEALQINDTRIMDAVTHLETFYNE-EKDKKFAVIE 642
 Db 253 KDLKANGRYONRIARSTSPRACYSALSLASCINVEHLELET----- 298
 QY 643 DSDDEGDDDEYCDGDEDEDLKKPLKIDETDRPLMTLFFENMKRLAEINPEY----- 696
 Db 299 -----QG-----IRPL-----HOYLLRLKKEKTKAAGLLAIIDPTRAMBL 334
 QY 697 -----ENKLTUKLRNTIMQYTTESARGIIFTKROSAYALSOQITENEKFAEY 747
 Db 335 TRRAMSGVHPKLDRLME-TLRRELKGDG-AHIVFTQFRDILLEYO-----CKKRE 386
 QY 748 GYVAHHLIGAGHSSEFKPTONEKEVYSKFTGKINLJATVVAERGLKESCNVIRY 807
 Db 387 GINAIVKFGGNSRSGEGLQKQRODILKSFRRGNHDIILSTVAEGITIPVLDLVAM 446
 QY 808 GLVTNETAMQAGRARADESTVYLAHSGSGVIEHETVNDPREKMKMYKTHCVQNNKPE 867
 Db 447 EPPVSEIRMIQRRGRTGRKRGKRMV-----LITEKTRD-----EAYYSSEIKRERSMK 496
 QY 868 EYAHKILQMOISIMEKMKTKKNIKHYNNPSLITFIKKNISVLAGSNDIDHVEKHI 927
 Db 497 NLGGGSVNVVNPIMEP-----SDEGPFYIADSK 525
 QY 928 HVN-MTPFEKELIYVENKALOKKADYOINCEII 961
 Db 526 EVNSRVLRRLKIKGVDELKPL--AVGDYQISDITI 554

RESULT 13

Db 84386 ATP-dependent RNA helicase homolog eIF-4A [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 10-May-2001
 C:Accession: D84386
 R:Wu, W.V.; Kennedy, S.P.; Mahalingam, G.G.; Bergquist, B.; Par, M.; Shukla, H.D.; Lasky, S.; Leitman, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Madlocks, D.G.; Jablo, J.; Jung, K.H.; Alm, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: D84386
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-784 <STO>
 A:Cross-references: GB:AE004437; NID:g10581766; PIDN:AG20456.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: eif4a
 C:Superfamily: ATP-dependent RNA helicase eIF-4A

Query Match 5.9% Score 313.5 DB 4 Length 784.
 Best Local Similarity 21.6% Pred. No. 1.5e-08;
 Matches 126; Conservative 91; Mismatches 190; Indels 175; Gaps 22;

QY 306 LOIRPOMEVAPALGKNNIICLPFGSKTRAVYIAADHLDKKKKASPEQVIVVVK 365
 Db 18 IEARVQLOLAAAROG-FTLVCLPTGLTKTVSLVTVYRL-----ADDAAGTALIAPT 72
 QY 366 VLLVEOLFREFOPLKKVY-----VIGSGDTOLKISFPEVYS-----CIHIIIS 412
 Db 73 KPLVEQ-----HAGTYRELAIPDDDYVVFYETR-----PDDRKAANTDAKVVYA 118
 QY 413 TAOILNLSNLNLENGEDAGVQSDFSLLIIDECHHTNKFAVYNNIMRHYLMOKLNRRK 472
 Db 119 TPQVENDLVG-----GRIDMDVYCHTFDECHRAVGSYVFLASNT-----IQMAR----- 164
 QY 473 NKPIVPLPOLGLTASGCGATKQAQAEHILKLCANLDAFTIKTKYKENDIQ 532
 Db 165 -----AAPLVYAMSASFG-----GTEAEIRTCENLGVNVEVTE----- 202
 QY 533 QEPCKKFAIADATREDPFREKLEIMTRIQTCCQSPMSDGTOTVEOMAIQMEK 569
 Db 203 -----DA-----DVGEHIDTDVOM----- 217
 QY 590 AKGNKERYCAEHLKRYNEALQINDTRIM--DAYTHLET-----YNEE 633
 Db 218 -----EHTVLP-----EILEVRDAINDVIEDRLAKREIGTKASSPDISUKIMEI 265
 QY 634 KDKKFAVIEDDSDGDDDEYCDGDEDEDLKKPLKIDETDR-LMTLFFEN----- 683
 Db 266 KARLOOLIDDSDGQGSVHAEVYK--LKRAVELVETQSVESRYRFEKRYANANIAG 323
 QY 684 -NMKLRLEANEYENKLT-----KLKNT--IMEQYRTESAKGIIITKIRGSA 731
 Db 324 ASASQRLVSEPRVAKMTAREFDGLHFKFQAMLLAETIGIDGDRIVITFSKITA 383
 QY 732 VALSOMITNEKFAEVYGAHHLIGAGHSSEFKPTONEKEVYSKFTGKINLJATV 791
 Db 384 EALTAFLEIGH-----FDRRFVQGDGSDGDMGTQTEQRETLAEFRNGDEHVLASTV 436
 QY 792 AEGGLDKKCNVIRYGLVTNETAMQAGRARADESTVY 833
 Db 437 AEBGLDPEVDLVLFEPVPTAIRSVQKRGKRGOTAGRNVV 478

RESULT 14

Db 548436 probable RNA helicase YIR002c - yeast (Saccharomyces cerevisiae)
 M:Alternate names: hypothetical protein YIB2c
 C:Species: Saccharomyces cerevisiae
 C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
 C:Accession: S48436; S50885
 R:Badcock, K.; Churcher, C.
 submitted to the EMBL Data Library, August 1994
 A:Reference number: S48432
 A:Accession: S48436
 A:Molecule type: DNA
 A:Residues: 1-993 <BAD>
 A:Cross-references: GB:Z47047; EMBL:Z38062; NID:g603997; PID:g764447; MIPS:YIR002c
 R:Voss, H.; Tanamas, J.; Teodoru, C.; Valencia, A.; Sensen, C.; Kilmann, S.; Schwaiger, V.
 Yeast 11, 61-78, 1995
 A:Title: Nucleotide sequence and analysis of the centromeric region of yeast chromosome
 A:Reference number: S50795; MUID:95282515; PMID:7762303
 A:Accession: S50885
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-993 <VOS>
 A:Cross-references: EMBL:X79743
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C:Genetics:
 A:Gene: SGD:MPI1
 A:Cross-references: SGD:S0001441
 A:Map position: 9R
 C:Keywords: ATP; nucleotide binding; P-loop
 F:107-114/Region: nucleotide-binding motif A (P-loop)
 F:205-210/Region: nucleotide-binding motif B
 F:209-212/Region: DEAD/H motif

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame-plus.p2n model

Run on: May 16, 2003, 10:02:54 ; Search time 65 Secs
(without alignments)
4836.057 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311
Sequence: 1 MSNCGSTDENFRYLISCFRA.....LPITPNLDVSHVCLFSDDEI 1025

Scoring table:

| | |
|---------------------------|--|
| BIOSUM62 | |
| Xgapop 10.0 , Xgapext 0.5 | |
| Ygapop 10.0 , Ygapext 0.5 | |
| Fgapop 6.0 , Fgapext 7.0 | |
| Delop 6.0 , Delext 7.0 | |

Searched: 441362 seqs, 153336381 residues

Total number of hits satisfying chosen parameters: 8827.4

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlh
-o=/cgn2_6/ptodata/1/ina/5A_COMB.seq:
-DB=Issued_Patents_NA -QFMT=fastcap -SUFFIX=mi -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bitsum62 -TRANS=trans40.cdi
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -VALLEN=2000000000
-USER=US09515363 @CGN.1.1.32.etrnat.16052003_100241.24719 -NUPU=6 -ICPU=3
-NO_XIPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7
```

Database :

```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|-------------------|
| 1 | 472 | 8.9 | 301 | 1 | US-08-143-576-7 |
| 2 | 335.5 | 6.3 | 7037 | 4 | US-09-853-768-3 |
| 3 | 227.5 | 4.3 | 193 | 1 | US-08-143-576-6 |
| 4 | 212 | 4.0 | 5852 | 4 | US-09-853-768-10 |
| 5 | 207.5 | 3.9 | 6773 | 4 | US-09-166-350-27 |
| 6 | 199.5 | 3.8 | 4868 | 5 | US-08-139-937-12 |
| 7 | 199.5 | 3.8 | 4868 | 5 | PCR-US93-11310-12 |
| 8 | 194.5 | 3.7 | 8789 | 1 | US-08-328-254-5 |
| 9 | 191.5 | 3.6 | 10136 | 1 | US-08-353-700-2 |
| 10 | 191.5 | 3.6 | 10136 | 5 | PCR-US95-16215-2 |
| 11 | 188.5 | 3.5 | 4363 | 2 | US-08-685-576-5 |
| 12 | 187.5 | 3.5 | 6921 | 4 | US-09-643-597-117 |

| 13 | 187 | 3.5 | 5020 | 4 | US-08-961-527-142 | Sequence 142, App |
|----|-------|-----|-------|---|---------------------|-------------------|
| 14 | 186.5 | 3.5 | 5361 | 4 <td>US-08-973-462-2</td> <td>Sequence 2, Appl</td> | US-08-973-462-2 | Sequence 2, Appl |
| 15 | 186.5 | 3.5 | 6152 | 1 <td>US-08-973-462-1</td> <td>Sequence 1, Appl</td> | US-08-973-462-1 | Sequence 1, Appl |
| 16 | 180.5 | 3.4 | 3393 | 4 <td>US-09-104-3248-1</td> <td>Sequence 1, Appl</td> | US-09-104-3248-1 | Sequence 1, Appl |
| 17 | 180.5 | 3.4 | 3393 | 4 <td>US-09-162-713-1</td> <td>Sequence 1, Appl</td> | US-09-162-713-1 | Sequence 1, Appl |
| 18 | 179 | 3.4 | 2001 | 4 <td>US-09-134-001C-2628</td> <td>Sequence 2628, Ap</td> | US-09-134-001C-2628 | Sequence 2628, Ap |
| 19 | 174 | 3.3 | 5053 | 2 <td>US-08-685-576-2</td> <td>Sequence 2, Appl</td> | US-08-685-576-2 | Sequence 2, Appl |
| 20 | 174 | 3.3 | 11958 | 4 <td>US-09-134-246-8</td> <td>Sequence 8, Appl</td> | US-09-134-246-8 | Sequence 8, Appl |
| 21 | 173 | 3.3 | 3636 | 4 <td>US-09-134-001C-1983</td> <td>Sequence 1983, Ap</td> | US-09-134-001C-1983 | Sequence 1983, Ap |
| 22 | 171.5 | 3.2 | 7105 | 4 <td>US-09-308-375-1</td> <td>Sequence 1, Appl</td> | US-09-308-375-1 | Sequence 1, Appl |
| 23 | 167 | 3.1 | 11091 | 4 <td>US-09-134-001C-2243</td> <td>Sequence 2243, Ap</td> | US-09-134-001C-2243 | Sequence 2243, Ap |
| 24 | 166.5 | 3.1 | 5661 | 4 <td>US-08-938-105-2</td> <td>Sequence 2, Appl</td> | US-08-938-105-2 | Sequence 2, Appl |
| 25 | 164 | 3.1 | 4739 | 3 <td>US-08-685-871-1</td> <td>Sequence 1, Appl</td> | US-08-685-871-1 | Sequence 1, Appl |
| 26 | 160 | 3.0 | 30549 | 4 <td>US-09-134-001C-322</td> <td>Sequence 322, App</td> | US-09-134-001C-322 | Sequence 322, App |
| 27 | 159.5 | 3.0 | 2871 | 4 <td>US-09-134-001C-1615</td> <td>Sequence 1615, Ap</td> | US-09-134-001C-1615 | Sequence 1615, Ap |
| 28 | 159 | 3.0 | 19307 | 3 <td>US-08-836-022A-10</td> <td>Sequence 10, Appl</td> | US-08-836-022A-10 | Sequence 10, Appl |
| 29 | 159 | 3.0 | 19307 | 4 <td>US-08-427-048A-10</td> <td>Sequence 10, Appl</td> | US-08-427-048A-10 | Sequence 10, Appl |
| 30 | 155.5 | 2.9 | 13425 | 4 <td>US-08-961-527-151</td> <td>Sequence 151, App</td> | US-08-961-527-151 | Sequence 151, App |
| 31 | 155 | 2.9 | 3825 | 4 <td>US-09-208-742-3</td> <td>Sequence 3, Appl</td> | US-09-208-742-3 | Sequence 3, Appl |
| 32 | 155 | 2.9 | 4582 | 2 <td>US-08-993-228-9</td> <td>Sequence 9, Appl</td> | US-08-993-228-9 | Sequence 9, Appl |
| 33 | 155 | 2.9 | 43676 | 3 <td>US-09-356-952-12</td> <td>Sequence 12, Appl</td> | US-09-356-952-12 | Sequence 12, Appl |
| 34 | 154.5 | 2.9 | 4248 | 3 <td>US-08-678-614-1</td> <td>Sequence 1, Appl</td> | US-08-678-614-1 | Sequence 1, Appl |
| 35 | 154 | 2.9 | 7808 | 4 <td>US-09-453-702B-247</td> <td>Sequence 247, App</td> | US-09-453-702B-247 | Sequence 247, App |
| 36 | 152 | 2.9 | 5093 | 1 <td>US-08-468-036-23</td> <td>Sequence 23, Appl</td> | US-08-468-036-23 | Sequence 23, Appl |
| 37 | 152 | 2.9 | 5093 | 2 <td>US-08-376-043-23</td> <td>Sequence 23, Appl</td> | US-08-376-043-23 | Sequence 23, Appl |
| 38 | 151 | 2.8 | 3741 | 4 <td>US-09-541-782-9</td> <td>Sequence 9, Appl</td> | US-09-541-782-9 | Sequence 9, Appl |
| 39 | 151 | 2.8 | 3741 | 4 <td>US-09-723-820-9</td> <td>Sequence 9, Appl</td> | US-09-723-820-9 | Sequence 9, Appl |
| 40 | 151 | 2.8 | 4308 | 4 <td>US-09-592-054-1</td> <td>Sequence 1, Appl</td> | US-09-592-054-1 | Sequence 1, Appl |
| 41 | 150.5 | 2.8 | 3107 | 4 <td>US-08-213-419B-1</td> <td>Sequence 1, Appl</td> | US-08-213-419B-1 | Sequence 1, Appl |
| 42 | 150.5 | 2.8 | 8948 | 4 <td>US-09-643-597-119</td> <td>Sequence 119, App</td> | US-09-643-597-119 | Sequence 119, App |
| 43 | 148.5 | 2.8 | 371 | 4 <td>US-09-222-575-46</td> <td>Sequence 46, Appl</td> | US-09-222-575-46 | Sequence 46, Appl |
| 44 | 148.5 | 2.8 | 1891 | 4 <td>US-08-973-462-3</td> <td>Sequence 3, Appl</td> | US-08-973-462-3 | Sequence 3, Appl |
| 45 | 148 | 2.8 | 3572 | 2 <td>US-08-713-815A-2</td> <td>Sequence 2, Appl</td> | US-08-713-815A-2 | Sequence 2, Appl |

ALIGNMENTS

```
RESULT 1
US-08-143-576-7
: Sequence 7, Application US/08143576
: Patent No. 5643761
:
: GENERAL INFORMATION:
: APPLICANT: Fisher, Paul B.
: APPLICANT: Jiang, Hongping
: TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRUCTURED
: TITLE OF INVENTION: CNA LIBRARY AND USES OF THE GENERATED LIBRARY
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John P. White, c/o Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10112
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/143,576
: FILING DATE: 25-OCT-1993
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 43563/JPW/ANC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
```

LENGTH: 301 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-143-576-7

Alignment Scores:
 Pred. No.: 5,32e-45 Length: 301
 Score: 472.00 Matches: 97
 Percent Similarity: 97.988 Conservative: 0
 Best Local Similarity: 97.988 Mismatches: 2
 Query Match: 8.898 Indels: 1
 DB: 1 Gaps: 0

US-09-515-363C-2 (1-1025) x US-08-143-576-7 (1-301)

OY 476 LysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyIly 495
 DB 3 AACCGAGTATCCCTCTCTCAGATACGAGCTAACACGCTTCACCTGCTGTTGACGCG 62
 OY 496 AlaThrLysGlnAlaLysAlaGluGluHisIleLeuLysLeuGlyAlaAsnLeuAspAla 515
 DB 63 GCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 122
 OY 516 PheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnIlePro 535
 DB 123 TTACGATTTAAACCTGTTAAAGAAACCTGATCAACAAACAAATACAGAGGACA- 181
 OY 536 CysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGlnLysLeuLeu 555
 DB 182 TGCAGAAAGTTTCCCTGATGACATGCACGACGACGACGACGACGACGACGACGACG 241
 OY 556 GluIleMetThrArgIleGlnThrTyrcysGlnMetSerProMetSerAspPheGly 574
 DB 242 GAAATTAATGACAGATTCACAACTTATGTCAATATGATCAATGTCAGATTGGA 298

RESULT 2

US-09-853-768-3
 Sequence 3, Application US/09853768
 Patent No. 6444466
 GENERAL INFORMATION:
 APPLICANT: Donna T. Ward
 APPLICANT: Andrew T. Watt
 TITLE OF INVENTION: ANTISENSE MODULATION OF HELICASE-MOI EXPRESSION
 FILE REFERENCE: RTS-0217
 CURRENT APPLICATION NUMBER: US/09/853,768
 CURRENT FILING DATE: 2001-05-10
 NUMBER OF SEQ ID NOS: 91
 SEQ ID NO 3
 LENGTH: 7037
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (183)...(5957)
 US-09-853-768-3

Alignment Scores:
 Pred. No.: 6.03e-27 Length: 7037
 Score: 335.50 Matches: 186
 Percent Similarity: 36.468 Conservative: 113
 Best Local Similarity: 22.688 Mismatches: 233
 Query Match: 6.328 Indels: 288
 DB: 4 Gaps: 43

US-09-515-363C-2 (1-1025) x US-09-853-768-3 (1-7037)

OY 309 ArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleGly 328
 DB 315 AGAAATATCAGCTGTAACCTGCTTGAAGCAGCTCTGAT--CATTAATCATCTGCTGT 371

OY 329 LeuProThrGlySerGlyLysThrArgValAla---ValTyrIleAlaLysAsp----- 345
 DB 372 TTAACACTGCGCTCAGGAGAGACATTATGCTAGTACTACTCTAAAGACGCTGCTC 431
 OY 346 HisLeuAspLysLysLysLysAlaSerGluProGlyLys---ValIleValLeuValAsn 364
 DB 432 TATCTAGATCTAGGGGAGCTTCAAGTAAATGAGAAAGACGCTTCTTCTTCTCAAT 491
 OY 365 LysVal---LeuLeuValGlnGlnLeuPheArgLysGluPheGlnProPheLysLys 383
 DB 492 TCTGCAACACGAGTGGCTTACACAGGTGCAGCTGTGACAACTCATTCAGATTCAG-- 548
 OY 384 TrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe----- 400
 DB 549 -----GTTGGGAAATACCAAAACCTAGAAATATGCATCTTGACCAAAAGAG 596
 OY 401 -----ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeu 417
 DB 597 AGATGGAACCAAGAGATTACTACTAAGCAC--CAGGTTCTCATTAATGACTTCAAGTCCG 653
 OY 418 GluAsnSerLeuAsnLeuAsnGluAsnGlyGluAspAlaGlyValGlnLeuSerAspPhe 437
 DB 654 TTGAATGTTTAAAAAT-----GTTACTTATCACTGTCAGACATT 695
 OY 438 SerLeuIleIleIleAspGluCysHisHisThrAsnLysGluAlaValTyrAsnAsnIle 457
 DB 696 AACCTTTGGTGTGTTGATGAGTGTCATCTGTGCAATCCAGACACCCCTATTCATATTT 755
 OY 458 MetArgHisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysPro 477
 DB 756 ATGAAG-----CTCTGTGAA----- 770
 OY 478 ValIlePro---LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyValAla 496
 DB 771 ATTGTCCATCATGTCCTCGCTGCTTTGGACCTAACTGCTTCATT---TTAAATGAGAA 827
 OY 497 ThrLysGlnAlaLysAlaGluGlnHisIleLeuLysCysAlaAsnLeuAspAlaPhe 516
 DB 828 TGGATCCAGAGATTGGAAGAAAGTTTCAGAAACTA----- 866
 OY 517 ThrIleLysThrValLysGluAsnLeuAspGln-----LeuLysAsn 530
 DB 867 ---GAGAAATTTCTTAAGAGTAATGCTGAACCTGCACTGCTGTCTTAAATGAG 923
 OY 531 GlnIleGlnGluProCysLysLysPheAlaIleAlaAsp-----AlaThrArg 546
 DB 924 TATACTTCTCAGCATGTGAG-----ATTGTGTGATTTGTGACCATTTATTCACAGA 977
 OY 547 GluAspProPheLysGluLysLeuLeuLysIle----- 557
 DB 978 AGTGGCTTTATGAAGAGCTGCTGATGATTAAGAAAGACCTAATTATTCATGAT 1027
 OY 558 -----MetThrArgIleGlnThrTyrcysGlnMetSerProMet 570
 DB 1038 TGTAAATATATCTGATCAATCAAAAGAAAGAGATTCTTAAATTGCAACAACTACTA 1097
 OY 571 SerAsp-----PheGlyThrGlnProTyrGlnGlnIleProValAlaIle 583
 DB 1098 TCAGACTGTGCGCGCTATTTGTTGTTCTGTGACCTCTGTTGACAGATTAAGTAAATGGA 1157
 OY 584 GlnMetGluLysLysAlaLysLys-----LysGlyAsnArgLysGlu----- 597
 DB 1158 ATGATGTAAGAGAACTACAGAAATATACAAACATGAGCAAGAGACTGTATACAAA 1217
 OY 598 -----ArgValCysAlaGlnHisLeu 604
 DB 1218 TTTTATTTGTTACAGACACTTCTCTAAGAAATATACATGCACTATGGAATATCTTC 1277
 OY 605 ArgLysTyrAsn-----GluAlaLeuGlnIleAsnAspThr 616
 DB 1278 TCACCTGCTCACTTGACCTGAATTTGTTAAGTCTTAAAGTAATCAACGCTGGAATTC 1337


```

QY 651 AspGluTyrCysAspGlyAspGluAspGluAspAspLeuLysProLysLeuAsp 6700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 GATATTATGTGTCATGAGTCAGTCGATTCCTGAGCATGAT-----CAT 56
QY 671 GluThrAspArgPheLeuMetThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeu 6900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 GAGCATGCAA----- 65
QY 691 AlaGluAsnProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGlu 7100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 -----GAAATTGAAAGAAAAAGAGAAAGCCACAGACAAATTTCCTCT 1070
QY 711 GluTyrThrArgThrGluGluSerAlaArgGlyIleLeuPheThrLysThrArgGluSer 7300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 CCTTTACCAACACT-----TTGTGGGCAATTATTTTGTGGAAATACAGTATACAA 1580
QY 731 AlaTyrAlaLeuSerGluIntPrIleThrGluAsnGluLys-----Phe 7440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 GCAGCTTGCTTAAACACAGATTGATTAAGAGAAAGCTGGCAACACAGATCCAGACCTGGCTTAT 2180
QY 745 AlaGluValGlyValLysAlaIleHisIleLeuIleGlyValaGluLys----- 7590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 ATCAGTAGCAATTTCCATTAAGTCAGCATGCGATTGGCAAGATAGCCCTCTAACCAACAAG 2780
QY 760 ---SerSerGluPheLysProMetThrGluAsnGluGluLysGluValIleSerLysPhe 7780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 ATCGAAGCAGCAATTCCACA-----AAACGAGAAAGAGGACTTACGAAATTT 3230
QY 779 ArgThrGlyLysIleAsnLeuLeuIleAlaThrThrValAlaGluGluGlyLeuAspIle 7980

```

| Alignment Scores: | | | |
|---|---------|--|---------|
| Pred. No.: | 2,5e-12 | | |
| Score: | 207.50 | | |
| Percent Similarity: | 36.73% | | |
| Best Local Similarity: | 18.03% | | |
| Query Match: | 3,91% | | |
| DB: | 4 | | |
| US-09-515-363C-2 (1-1025) x US-09-166-350-27 (1-6773) | | | |
| QY | 41 | GLVALVYSGIUGLIGLEGINARGTHVALAATHRSERGLYASMETGINALVAVGLU | 60 |
| DB | 251 | AAATGATGACGACGAGTTGACGATTCGTATACCAAGATGGACAGTGCACAT | 301 |
| QY | 61 | LEULEULESERTHRLEUGLULYSGLVATTPHLSLEUGLYTPRTHRGGLINLVAL | 80 |
| DB | 302 | -----AAGGAGTTGGAAACATCA-----CATTTAAACATGTGGAAGAA | ATT 343 |
| QY | 81 | GLVALALEUARGARGTHRGILYSERPROLEUALALAAIRGYR | 94 |
| DB | 344 | GAATATTTGAAAAAATGAGTTGATGCGCGATGCTGCATTCGCAAAATACAGTGAAGACAAATTAAC | 403 |
| QY | 95 | -----MetASpPROGLULEUTHRASPLEUPROSPROSER | 106 |
| DB | 404 | TTACAAAGCAGCTGAGAGAACGACATGAATACCGAATTAACACTTCAGAAATTAATAA | 463 |
| QY | 107 | PHGLIUSMPLAHSASPGLUTYRLEUGINLEULEU-----ASNTPLAUGLIN | 122 |

Db 464 TTTCAGACAACTCTGACATATGTTAAAACTACAGAGAGATTGCAAAATTAG 523
 Oy 123 ProThrLeuValAspLysLeuLeu---ValArgAspValLeuAspLysCysMetGluGlu 141
 Db 524 CCAGGCTTTGAGGAGCAAAATTTTATATCTGCAAAAGCAATTAGACCTACTACTGTATA 583
 Oy 142 GluLeuLeuThrIleGluAspArgAsnArgIleAlaIleAlaGluAsnAsnLysAsnGlu 161
 Db 584 AAGAACGAAACAGATTCTCACTCCCAAAATATCATGTAGAGCTTAATCTCCAGATTAGCA 643
 Oy 162 SerGlyValArgGluLeuLeuLysArgIleValGlnLysGlnAsnTrpPheSerAlaIle 181
 Db 644 AAAAATATTATTAAGTTTGGAGAGAGCTTTTACAG----- 679
 Oy 182 LeuAsnValLeuArgIleThrGlyAsnAsnGluLeuValGlnGluLeuThrGlySerAsp 201
 Db 680 TTGAAGCTATACACCAAGACAGCTGAAAGAGATTGATGTCGACGATTAAGCATCA--- 736
 Oy 202 CysSerGluSerAsnIleGluIleGluAsnLeuSerGluValAspGlyProGluValGln 221
 Db 737 GCTAAGCAACATGACACAGATTAATTAAGTTGAACAGATTAAGACAGAACTTAGTAATA 796
 Oy 222 Glu---GlnLeuLeuSerThrValGlnProAsnLeuLeuLysGluValTrpGlyMet 240
 Db 797 CATGTGAGGCAAGTGAAGAAACATCCAGAGAAATATCAATGTGCTTAAGAAATTA 856
 Oy 241 GluAsnAsnSerSer-----Glu 246
 Db 857 AGCAAAAGCCACCTCAATCCAAACCAAGACATCATATGTTCTATCTCTTCGCAAGAA 916
 Oy 247 SerSerPheAlaAspSerSerValValSerGluSerAspTrpSerLeu----- 262
 Db 917 AATACATTTTGTA---CAAGTAGTAATGAAGAAAGTCACACACTTAGACATACCTTA 973
 Oy 263 ---AlaGluGlySerValSerCysLeuAspGluSerLeuLysLysAsnSerAsnMet 280
 Db 974 AAGAACTTGATCTCAACACATCTATTAAGATGAGTAGTAATTAATTAATTAATCTT 1033
 Oy 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaAlaArgAla 300
 Db 1034 ---AAGTTAAACTTGAAATGATGCTCAACATATTAAGATGAGTTT 1078
 Oy 301 SerProGluProGlu---LeuGlnLeuArgProGlyGlnMet----- 313
 Db 1079 TTTCATGAACGGAGACCTTAGAGTTAAATTAATGAATATATCTAGCTAAAGACAA 1138
 Oy 314 ---GluValAlaGlnProAlaLeuGluGluGlyLysAsnIleIleLys 328
 Db 1139 CAGGCTGTGTAATGTAATAATTAATCTGAGCTAGCAGAGTTAAATAACAGTTTGC 1198
 Oy 329 LeuProThrGlySerGlySerThrArgValAlaValTyrIleAlaLysAspHisLeuAsp 348
 Db 1199 ---TATACCTTAGAACACATTAACATA 1222
 Oy 349 LysLysLysLysLysLysLysLysLysLysLysValIleValLeuValAsnLysValLeu 367
 Db 1223 GAAGTACAGAGCTTTAAGGACAAACATCAAAAGAAATATCAAAATTAAGACAAATTT 1282
 Oy 368 LeuValAlaGluGlnLeuPheArgLysGluPheGlnProPheLeuLysLysTrpTyrArgVal 387
 Db 1283 TTGTGCAATTCAGAAAAAAGAAAA---TTAACATTAAATGTTGAATA 1327
 Oy 388 IleGlyLeuSerGlyAspThrGln---LeuLysIleSerPheProGluValValLysSer 406
 Db 1328 CAGGCTCTTAAGCAACAGTGTGAAACCTACACCAAGAAAGCAAGCAATTTAAAT 1387
 Oy 407 CysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSerLeuLeuAsnLeuGluAsn 426
 Db 1388 TATGAG-----AGTTTACAGAGATTTATGAA---ATTTCACAAACAGAGACTG 1432
 Oy 427 GlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLeuAspGluCysHis 446
 Db 1433 GGGGAATCTGCTGGAAAAAATTAAGTCAAGAGTTT----- 1465

Oy 447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 Db 1466 GAATCAATGAACACACACACACACATCTATCTT-----CATGAAGCTCAGACAGATCTC 1519
 Oy 467 LysAsn-----AsnArgLeuLysLysLys 474
 Db 1520 AAGACTGCTTTTACTGAAAAAGATGCCCTTCGCAAACTGGAATCGGCTCCAGAGAA 1579
 Oy 475 AsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGluValGly 494
 Db 1580 AATGAAGAGTTACTATCTCAACACCAAGAAATGGA-----CCAGAAATTGA 1624
 Oy 495 GlyAlaThrLys-----GlnAlaLysAlaGluGluHisIleLeuLysLeuCysAlaLys 512
 Db 1625 AATACCATTAAGAACCTTCACAGAAAGAAATGAGATGATCTACTACTAGCTC---AGTTAA 1681
 Oy 513 LeuAspAlaPhe-----ThrIleLysThrValLysGluAsnLeuAsp 526
 Db 1682 ACAGATACCATGTATAAAGAAATTAGACAGAAAGATAATTTCTCTACTAGAGAAATAAT 1741
 Oy 527 GlnLeuLysAsnGlnIleGlnGluProCysLysLysPheAlaIleAlaAspAlaThrArg 546
 Db 1742 GATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1780
 Oy 547 GluAspProPheLysGluLys-----LeuLeuGlnLeuMetThr 559
 Db 1781 ---GATTAATTTCCATTAAGCAAAATGTCACAAAGCAAGAACTGATTTGAACTTCTAG 1837
 Oy 560 ArgIleGlnThrTyrCysGlnMetSer----- 568
 Db 1838 AAGATAGAGCAAAACATTCACAGTACACAGTACAGAAACAAAGATAATGAATTACAA 1897
 Oy 569 ---PrometSerAspPheGlyThrGlnProTyrGluGluTrp 581
 Db 1898 GGAGACTAGACAGACATTTAAAGAAAGATCAAAATTAAGCAAAACCTGAAAGAACTT 1957
 Oy 582 AlaIleGlnMet----- 585
 Db 1958 ATGTTCAATTAAGATCTCTCTGACACAAAGAAATTTCTGCTGAACTGAAAGTCT 2017
 Oy 586 ---GluLysLysAlaAlaLysLys----- 592
 Db 2018 CTTATGAGAAACATTAATTAATCACTGATCAGAAAAAACAAGTGTAGAGGATTTGAG 2077
 Oy 593 ---GlyAsnArgLysGluArgValCysAlaGluHisLeuArgLysTyrAsnGlu 609
 Db 2078 GTTTTCTGTCACAAAGAAAGATGTATCTTAAAGAAACATTAATTAATTAATTAAT 2137
 Oy 610 AlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPhe 629
 Db 2138 AAATCTTCACTTA-----ATGTTGAAGCAAGATTAATTAATTAATTAATTAAT 2182
 Oy 630 TyrAsnGluGluLysAspLysPheAlaVal----- 640
 Db 2183 CTTGAAATAGACAAAGTTCAGAAATTTTAAACAGTGTATAGTTTCTTAA 2242
 Oy 641 ---IleGluAspAspSerAspGluGly----- 649
 Db 2243 GAATGCGATCAGAAAGTTTCAGACAGAGTGAAGAAAGATGTTTAATGTCACAG 2302
 Oy 650 ---AspAspGluTyrCysAsp-----Gly 656
 Db 2303 GCAGTCGGTAATCTTGCAAAAAATTAATGAGAAATTAATGCTGCTTTCAAGCT 2362
 Oy 657 AspGluAspGluAspAspLysLysProLeuLysLeuAspGluThrAspArgPhe--- 675
 Db 2363 GATGAAAAAGTATGAGTTAGAAAAAGATTAAGTGCCTTCACAAAGAGTATAT 2422
 Oy 676 ---LeuMetThrLeuPhe-----PheGluAsnAsnLysMetLeuLysArg 689
 Db 2423 CAGTGTGAAGAACTTAAGCTTTATTGAGAGACTATGAGCAAGAAAGTCTCTTAAG 2482

QY 690 LeuAlaGluAsnProGluTyrGluAsnGluLys-----LeuThrLys 703
 Db 2483 ---AAAGAGTACAGAAAGAAATACAGTGAAGAGCCCTGAGTGTGATCTTCTAGAA 2539
 QY 704 LeuArgAsnThrIleMetGluGlnTyrThrArgThrGluGlnGlnSerAlaIleGlyIleIle 723
 Db 2540 ATGACAGATGCT-----ATGAAAAACAGAGCTTGAATTCAGAT-----CTT 2584
 QY 724 PheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLys 743
 Db 2585 TTATTCAGAGTGAAGAGATGCTCAACATGACGAAAAAGTGAATTCATATGAAAA 2644
 QY 744 PheAlaGluValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerGlyLeuPhe 763
 Db 2645 ---GAAAAATGTTTATTAAGAACAT-----GAAAACTA 2677
 QY 764 LysProMetThrGlnAsnGluGln-----Lys 772
 Db 2678 AAGCACTACTAGACAAAAAGAAATACGATAGCAGAGACAGACAGTTCATATTAAG 2737
 QY 773 GluValIleSerLysPheArgThrGlyLysIleAsnLeuLeuIleAlaThrValAla 792
 Db 2738 GATTCCTTACGAAAAATCACCCTTCTGTAAAAATGATCTCTCTCAGTAAAGAGCTG 2797
 QY 793 GluGluGlyLeuAsp-----IleLysGluCysAsnIleValIleArgTyrGlyLeuVal 810
 Db 2798 GAAGAAAAATACAAAAATCTCGAAAAAGATGCAAA-----GAAACGAGAGGAAA 2848
 QY 811 ThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspIleSerThrTyr 830
 Db 2849 ATAAATATAGATAAATATAGTGTGCGTAAAGGCAAAAGAACTATTCACGACAAA 2908
 QY 831 ValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAspPheArg 850
 Db 2909 -----GAGACCACAGCTGTGAAGCAAGAACTT 2935
 QY 851 GluLysMetMetTyrLysAlaIleHisCysValGlnAsnMetLys----- 865
 Db 2936 GAATCTCTTGATCAGAAAAAGACCACTTCTGCTTCATGAGACATCTATTACAGA 2995
 QY 866 ProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetGlnSerIleMetGluLysLys 885
 Db 2996 GCAGAAAGCTATAGAACTCTTATTAGAAATATGAAAGACTCA-----GAGCAA 3046
 QY 886 MetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAspProSerLysIleThrPhe 905
 Db 3047 CTGGATCTGGAAGAAAGACCTGCTAATATTGACCATCTGATTAAGAGCTTACAGAA 3106
 QY 906 LeuCysLysAsn-----CysSerValLeuAlaCysSerGlyGluAspIleHis 921
 Db 3107 CAATTAGAAATTCGACTTTCGAGTGTGAACCAATATTCATATGACGATCTC--- 3163
 QY 922 ValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluTyrIleVal 941
 Db 3164 ---CTGGCTCTGATTCAGACATACAGTCTAATGCCAAATTATTAAGATGATGTTTA 3220
 QY 942 ArgGluAsnLysAlaLeuGlnLysCysAlaAspTyrGlnIleAsnGlyGluIleIle 961
 Db 3221 GAAGTCAGAGACCC-----AAAGCAATGCTAGCAAGAAATAGAGTGAAGAACTT 3274
 QY 962 CysLysCysGlyLysIleAlaTrpGlyThrMetMetValHisLysGlyLeuAspLeuPro 981
 Db 3275 CAGAAAGACAGAGATA-----AAGAACATCTACTACTGTA 3313
 QY 981 IleuLysIleArgAsnPheValAlaValPhe-----LysAsnAspSerThrLysLysGlu 999
 Db 3314 AATGAACCTTCAACACTCTGACACACTTCAAAAAAGAAACAGACCTTACAGAAAGC 3373
 QY 999 nTyrLys-----LysTrpValGluLeuProIleThrPheProAsn 1012
 Db 3374 ATCGAAGAAATTAAGCTGGTTAA-----ACGATGCTGAC 3410
 RESULT 6

US-08-139-937-12
 Sequence 12, Application US/08139937
 Patent No. 5821070
 GENERAL INFORMATION:
 APPLICANT: LEE, MEN-HWA
 APPLICANT: SHAN, BEI
 TITLE OF INVENTION: CELLULAR GENES ENCODING
 TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CAMPBELL AND FLORES
 STREET: 4370 LA JOLLA VILLAGE DRIVE
 CITY: SAN DIEGO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/139,937
 FILING DATE: 20-OCT-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/979,156
 FILING DATE: 20-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: CAMPBELL, CATRYN
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 9370
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-9901
 TELEFAX: 619-535-8949
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4868 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-139-937-12
 Alignment Scores:
 Pred. No.: 1,14e-11 Length: 4868
 Score: 199.50 Matches: 219
 Percent Similarity: 35.52% Conservative: 187
 Best Local Similarity: 19.16% Mismatches: 472
 Query Match: 3,768 Indels: 265
 DB: 1 Gaps: 46
 US-09-515-363c-2 (1-1025) x US-08-139-937-12 (1-4868)
 QY 21 ArgValLysMetTyrIleGluValGluProValLeu----- 32
 Db 207 AAAATTTGAAGCATGATGAATTTGAAAAAATGTTGGGAACTTAAGAAAGAAATTTCA 266
 QY 33 -----AspTyrLeuThrPheLeuProAlaGluValLysGluGlnIleGlnArgThr 49
 Db 267 GATTATAGTGAAGAAATTTGAAATATTTTCTGTGTATCAACAGAGATTTCTGACGACTA 326
 QY 50 ValAlaThrSerGly-----AsnMetGlnAlaValGluLeuLeuLeuSer 64
 Db 327 GAACCTTTGAAGGCCCTCAATCTGATTAGAAATATGCATGACAGATAAATCATACAGTGA 386
 QY 65 ThrLeuGluLysGlyValIleThrHisLeu-----GlyTrpThrArgGluPheValGluAla 82
 Db 387 GATATTGAGATAAATGTCGCCAAGCTGATGACACAGCTGAAGAGACGATTTCTTAT 443
 QY 83 LeuArgArgThrGlySerProLeuAlaIleAlaArgTyrMetAsnProGluLeuThrAspLeu 102
 Db 444 -----GTGAAATGAGCTGAGTAGATC 467


```

Db 2499 GTGCTGAGAGTCTTCAAGAACTCTAGACATAGAC-----CTTCATTTGACAAAA 2549
OY 727 ThArgInserAlaTyrAlaLeuSerGlnTrpIleThrGlnAsnGlyLysPheAlaGlu 746
Db 2550 ATGACACAAAATGCTCTTTGTT-----CAGAAAGTAACAAA 2585
OY 747 ValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGlnPheLysPromet 766
Db 2586 ATGCTCTGCAAGAAAGAACTGAGCTGCAGAGGAAATGCGATGATGCGACAGAAACAGCA 2645
OY 767 ThrGlnAsnGlnGlnLysGlnValIleSerLysPheArgThrGlyLysIleAsnLeuLeu 786
Db 2646 GAGCTGCAACAGACAGCTCATGAGAGAAAATAGCTAGC TCGAGAGTTCAGCTTACTG 2705
OY 787 -----IleAlaThrVal 791
Db 2706 TTGGAGAAATTAAGACAGCAAGATCAATGAGAGCTCACACTAGAAATAGTCAA 2765
OY 792 AlaGlnGlnLysLeuAspIleLysGlnLysAsnIleValIleArgTyrGlyLeuValThr 811
Db 2766 TTGAGAGAGGCTAGCTAGCTGACACAAAGACAGCTGAAAGAGAGAGAGAGAGAGA 2825
OY 812 AsnGlnIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspLysSerThrTyrVal 831
Db 2826 GAGGAATAGCTGAAATATCATGCTACGCTTCATGAAAGCTGAAAGAAACAGAGGCTTTG 2885
OY 832 LeuValAlaHisSerGlySerGlyValIleGlnHisGlnThrValAsnAspPheArgGlu 851
Db 2886 CTTTGGACACAAACAAAGATGAGAAATCCAGCA-----TACCGACAG 2936
OY 852 LysMetMetLysLysAlaIleHisCysValGlnAsnMetLysProGlu----- 867
Db 2937 AAATGACTCTTCTAAGAA-----GAATGCTCAGTTCACAGAGCTGAGATAGACCTTTA 2993
OY 868 -----GlnTyrAlaHisLysIleLeuGlnLeuGlnMetGlnSerIleMetGln 883
Db 2994 AAGCTAGTAAAGAGAGCTCATTAATCTATTGAAAGCTACTACTAG-----ATTGGAAA 3050
OY 884 LysLysMetLysThrLys-----ArgAsnIleAlaLysHisLysTyr 896
Db 3051 GAATTCAGAGAAACCAAGATGACAACTTAATATGTAAATGATGAAGCAAGAAAT 3110
OY 897 LysAsnAsnProSerLeuIleThrPheLeuGlyLysAsnLysSerValIleAlaLysSer 916
Db 3111 GAACGTGCCAGGAGAAATGAAGTTGTGATCAAAATCTGTAAACGCTGAAA----- 3164
OY 917 GlyGluAspIleHisValIleGlnLys----- 925
Db 3165 ---GAGGAAAAGAGATATGCGAGAAAGAACTCTCAATTCACAGCTGACAGAGAG 3221
OY 926 -----MetHisHisValAsnMetThrProGlnPheLysGln 937
Db 3222 CAGAAAACAGTACTGTATGATACCAAGCTGATGATTAACAACTGACATCAAGAAA 3281
OY 938 LeuTyrIleValArgGlnAsnLysAla-----LeuGlnLysLysCys 951
Db 3282 CTAAAGAAAGCTTGAAGAAAACCAAGAGAGAGAGATGATGATGATGAAGTACTGT 3341
OY 952 AlaAspTyrGlnIleAsnGlnGlnIleIleCysLysCysGlyGlnAlaTrpGlyThrMet 971
Db 3342 TCC---TTGCTTATAGCATGAAAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACA 3398
OY 972 MetValHis 974
Db 3399 GTGGCCCAT 3407

```

```

TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11310
FILING DATE: 19-NOV-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHERYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: FP-CJ 9790
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4868 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US93-11310-12

Alignment Scores:
Pred. No.: 1,14e-11 Length: 4868
Score: 199.50 Matches: 219
Percent Similarity: 35.52% Conservative: 187
Best Local Similarity: 19.16% Mismatches: 472
Query Match: 3,768 Indels: 265
DB: Gaps: 46

US-09-515-363c-2 (1-1025) x PCT-US93-11310-12 (1-4868)
OY 21 ArgValLysMetLysIleGlnValGlnProValLeu----- 32
Db 207 AAATGGAAGCATGATGAAATGGAATAATGTTGGGAACTTAAGAAAGAAATCA 266
OY 33 -----AspTyrLeuThrPheLeuProAlaGlnLysGlnGlnIleGlnArgThr 49
Db 267 GATTTAAGTGAATAATTTGAAATATTTCTTGTGTACACAGAGATTCTCCAGAGCTA 326
OY 50 ValAlaThrSerGly-----AsnMetGlnAlaValGlnLeuLeuSer 64
Db 327 GAAACTTCTGAAGGCTCAATTCATTTAGAAATGATGCGATGATCAATCACTACGTTAA 386
OY 65 ThrLeuGlnLysGlyValTyrHisLeu-----GlyTyrThrArgGlnPheValGlnAla 82
Db 387 GATATTGAGATAATATGTGCCAAGGTGAATGACACCTGGAAGAGAGATTTCTTGAT--- 443
OY 83 LeuArgArgThrGlySerProLeuAlaIleArgTyrMetAsnProGlnLeuThrAspLeu 102
Db 444 -----GTGAAAAATGAGCTGATGATGATC 467
OY 103 ProSerProSerPheGlnAsnAlaHisAspGlnTyrLeuGlnLeuLeuAsnLeuLeuGln 122
Db 468 AGATCGAGAAAGCTAGCATTTGAGCAT---GAGAGCTCTACTCTCGAGAGCTGACTTACAG 524
OY 123 ProThrLeuValAspLysLeu---LeuValArgAspValLeuAspLys----- 137
Db 525 GTAGTTCAACAGACAGCACTATGTTTACAAAAGACATGAATAATTAAGACAGCTTATTT 584
OY 138 ---CysMetGlnGlnGlnLeuLeuThrIle---GlnAspArgAsnArgIleValAlaAla 155

```

```
|||||
585 GTCGCTGTAAGAAAGAACTCTCAGTGTGACAAAGTGAAGAAACAGATCGTGCAGAA 644
156 GUAASnArgLeuLysGlu-----SerGlyValArgGluLeuLysArgValLeuGln 173
645 TTGATACTACTATGTCAAAAAACCACGCACTGATCATGTTCTCTCAAAAAATGAAGAG 704
174 Lys-----GluAsnTrpPheSerAlaPheLeuAsnValLeuArgInLys 188
705 AAAACACAGAGCTGATGATCTCATCAAAAGTATGTCATGTCATTCACAGTGCAGAG 764
189 -----GlyAsnAsnGluLeuValGlnGluLeuPheGlySerAspGlySerGln 204
765 GCAGAGTGAAGGAAAAAGACGAACTCTCAGACCTTGGCC---TCTGATGAGCTGAG 821
205 SerAsnAlaGluLeuGluLeuLeuSerGln---ValAspGlyProGluValGlnGln 223
822 CTGTTAAAGACAAACACTATCTCCAGAAAGCTGCAGAGTTTGAAAAAGCTCAAG 881
224 LeuLeuSerThrThr-----ValGlnProAsnLeuGlnLys 245
882 GCACCTGCTTGCACAAATGTGAGTGAAGAAACCAATTGCACTGAAATGAAGAGAA 941
236 GluValTrpGlyMetGluAsnAsnSer-----SerGlySerPheAla 250
942 GAATTCCTGTCAGAGAACTCTGAAGCCTGCAGCCAGATGAGATTCGATATATCA 1001
251 AspSerSerValAlaSerGluSerAspThrSerLeuAlaGluLys 266
1002 AACCTGAATGTCTCCAAAGCCCTTGAGGCCGACCTGTGTGAGAAAGTGCATTCG 1061
267 ---ValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMetGlySerAspGly 285
1062 AGCTGAGCTCAGACAGACAGAGGAGAGTG---CATCAGCTGAGAGAGGCAATGAG 1118
286 ThrMetGlySerAspSerArgGlu-----GluAsnValAlaAla 298
1119 AGAGTTCCGATTGAGCCGATGAAGAAAGACCTGCACATCCAGCAAGAACTGAAGA 1178
299 ArgAlaSerProGluProGluLeuGlnLeuArgProGlyMetGluValAlaGlnPro 318
1179 CGCGAGCGGAGAAATGATTAATTAAGTTGAGAACCTTGAAAGAAATTCGACG 1238
319 AlaLeuGluLysAsnIleIleCysLeuProThrLysGlyLysThrArgVal 338
1239 ATGTCAAGAAAGAAACCGAGAGTAGTATCTTGATGCCAGAAATCCAAAGCAGAA 1298
339 AlaValTrpIleAlaLys---AspHisLeuAspLysLysLysAlaSerGluProGly 357
1299 GAGACTTAAACACAAATAGAAAGATGCGCAGAGCCTGAAAGTTTGAATTAGAC 1358
358 LysValIleValLeuValAsnLysValLeuLeuValGlnLeuPheAlaGlyGluPhe 377
1359 CTGTTCAGTTAAGCTGTGAAGAAAGAAATCTCACAACAAATACAAAGAAACAGGT 1418
378 GlnProPheLeuLysLysTrpArgValIleGly-----LeuSerGly 392
1419 CAG-----TTTCAGAGAAAGACAGCACTACTCTCTCATTAAGTCTCTTGAAGAA 1472
393 AspThrGlnLeuLysIleSerPheProGluValAlaLysSerLysAspIleIleLeuSer 412
1473 AAGGAGACAGACAGATACAGATCAAAAGAAATCTAAA-----ACT 1514
413 ThrAlaGlnIleLeuGluAsnSerLeuLeuAsnLeuLys 426
1515 GCAGTGAAGATGCTTCAAGATCAAGTAAAGAGCTAAATAGACAGTAGAGCTTGT 1574
427 GlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLeuAspGlyCysHis 446
1575 GGTGACCAAGAAATTAATGAGGCGCAGACAGAGAGTCTACACCAAlAGAGAGAG 1634
447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeu 466
|||||
1635 CATCAGCTGAGAAATAGCATT-----GAAAAGCTG 1664
467 LysAsnAsnArgLeuLysLysGluAsnLysProValIleProLeuProGlnIle----- 484
1665 AAGAGCCCGCTGAGAGCTGATGAAAGAAAGCAGCTCTGTGTTACCAACATGAAAGAA 1724
484 ----- 484
1725 AGTGAGCATATGACAGATTACTTAAGGAGTAGAGTGAGAACTTGAAAGAGCTAGAG 1784
485 LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLysAlaGlu 504
1785 ATAGCCAGGACCAACCAAGACATCAGCTCTTGAGGACAGAGAAATTCAAAAGAGAGTA 1844
505 HisIleLeuLysLeu-----CysAlaAsnLeuAsp 514
1845 GAGACCTTAAAGCAAAATAGAGGAGTGAACCCAAAGTGCAGAGTCTGCAATTAAT 1904
515 AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 534
1905 GTTGTACTATATAGGTCAGAAAAAGAAATCTGACAAATGAAATTACAAAAGCAGAG 1964
535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheGlyGluLys 554
1965 CGAATATCTGAATTAGCAAAATTAATTCATCATTTGAAAAATTTTCCAAAGAAAA 2021
555 LeuGluIleMetThrArgIleGlnThrTyrCysGlnMetSerProMetSerAspPheGly 574
2022 -----GAGCAAGGAAAGTACAGATTAAGAAATTAAGAAATTAAGCAACTGCG 2063
575 ThrGlnProTrpGluGlnTrpAlaIleGlnMetGluLysLysAlaAlaLysLysLys 594
2064 ATGCAATCTCTTCAACACAAATTAAGAGCTCAATGAGAGGTGAGCCCTGATAT 2123
595 ArgLysGluArgValCysAla-----GluHisLeuArgLysTyrAsnGluAlaGln 612
2124 GACCAAGAAAGCCGTGAAGCCCAAGACGAAATCTAGTACAGTACAGTCTTGA 2183
613 IleAsn-----AspThrIleArgMetIleAspAla-----TyrThrHisLeu 626
2184 CTTGACAAAGCTCAGTGTGTACAGAGCCTTGATGAGCCCAAAATTAATATATCTT 2243
627 GluThrPheTyrAsn-----GluGlu 633
2244 CAATCTTCATGATGATGCTCATTCAGAAAGTAGAAGTGGCAGAGCAAACTGAGAG 2303
634 LysAspLysLysPheAlaValIleGluAspAspSerAspGluGlyLysAspProLys 653
2304 AAGGATGAAGAAATCAGTACAGTGAATAATCAAT----- 2339
654 CysAspGlyAspGluAspGluAspAspLysLysProLeuLysLeuAspGluProAsp 673
2340 -----CAAGACCAAGAGCAGCTTGTCTAAACTGTCCAGGTGAGAAATAGAG 2387
674 ArgPheLeuMetThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaAsn 693
2388 CAC-----CACTTGGAGAGAGCAAAACTTGAAGTGAAGATCTGACAGAGAG 2438
694 ProGluTyr-----GluAsnGluLysLeuThrLysLeuArgAsn 706
2439 TTGAGCAGAGATCCAGATGCTACAAATCCAAAAAGCTCTTGGAGACATTAAGAA 2498
707 ThrIleMetGluGlnTyrThrArgTrpGlnGluGluSerAlaArgGlyIleIlePheGlnLys 726
2499 GTGCTCAGAGTCTTACAAAGATCTAGAGAAATGAG-----CTTCATTTAAAGAA 2549
727 ThrArgGlnSerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLysPheAlaGlu 746
2550 ATGCACAAAAATGCTTGT-----GAAAAAGTAAAGAA 2585
747 ValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGluPheLysProMet 766
2586 ATGACTGCAAAGAAAGTGAAGTGCAGAGGAAATGATGATGACAGAGAAAGAGCA 2645
```


[illegible]

RESULT 9
 US-08-353-700-2
 : Sequence 2, Application US/08353700
 : Patent No. 5599919
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: YEN, TIMOTHY J.
 : APPLICANT: RATNER, JEROME B.
 : TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 : TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETICALLY RESPONSIVE PROTEIN
 : TITLE OF INVENTION: AND METHODS OF USE
 :
 : NUMBER OF SEQUENCES: 4
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: DANN, DOFFMAN, HERRELL AND SKILLMAN
 : STREET: 1601 MARKET STREET, SUITE 720
 : CITY: PHILADELPHIA
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103-2307
 :
 : COMPUTER READABLE FORM:

```

1  MEDIUM TYPE: Floppy disk
2  COMPUTER: IBM PC compatible
3  OPERATING SYSTEM: PC-DOS/MS-DOS
4  SOFTWARE: PatentIn release #1.0, Version #1.25
5
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/353,700
8  FILING DATE: 09-DEC-1994
9  CLASSIFICATION: 435
10 ATTORNEY/AGENT INFORMATION:
11 NAME: REED, JANET E.
12 REGISTRATION NUMBER: 36,252
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (215) 563-4100
15 TELEFAX: (215) 563-4044
16 INFORMATION FOR SEQ ID NO: 2:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 10136 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: double
21 TOPOLOGY: linear
22 MOLECULE TYPE: cDNA to mRNA
23 HYPOTHETICAL: NO
24 ANTI-SENSE: NO
25 ORIGINAL SOURCE:
26 ORGANISM: HUMAN
27
28 US-08-353-700-2

```

| | |
|------------------------|----------|
| Alignment Scores: | |
| Pred. NO.: | 3,47e-10 |
| Score: | 191.50 |
| Percent Similarity: | 35.60% |
| Best Local Similarity: | 19.26% |
| Query Match: | 3.61% |
| DB: | 1 |
| | Gaps: |
| | 45 |

| | | | |
|----|------|--|------|
| QY | 21 | ArgAllylsMetYLrIleGluValGlnProValLeu----- | 32 |
| | | :::::::::::::::::::: | |
| Db | 5919 | AAATTCAGACCATGCATAGCAATTGGAAAAAATAGTTGGGAACCTTACAGAAACAAGAAACTCA | 5978 |
| QY | 33 | -----AspTrpLeuThrPheLeuProAlaGluValLysGluGlnIleGlnValThr | 49 |
| Db | 5979 | GATTTAAGTGCAAAAATATTGCAATATTTTTCTGTGATCACCAGAGAGTTACTCATCAATACTA | 6038 |
| QY | 50 | ValAlaThrSerGly-----AsnMetGlnAlaValGlnLeuLeuProSer | 64 |
| | | ::::: | |
| Db | 6039 | GAACCTTCAGAGCCCTCAATTCTGATTAGAAGATGCAGATAAATCATCAATCTGA | 6098 |
| QY | 65 | ThreudGluLysGlyValITrpHisLeu-----GlyTrpThrargGluPheValIleAla | 82 |
| | | :::::::::: | |
| Db | 6099 | GATATTCAGACATTAATGGGCCAAGGTGAATGCACGCTGAGAGAGAGATCTTTAT--- | 6155 |
| QY | 83 | LeuArfArgThrGlySerProLeuAlaAlaArgTyMetAsnProGlnLeuThrAspLeu | 102 |
| Db | 6156 | -----GTGGAAAAATGAGCTGACTGATACATC | 6179 |
| QY | 103 | ProSerProSerPheGluAsnAlaHisAspGluTYrLeuGlnLeuLeuAsnLeuGln | 122 |
| | | :: | |
| Db | 6180 | AGATCGAGAACAGCTAGCATTTGAGCAT---GAAGCCCTCTACCTGGAGGCTGATCTTAGG | 6236 |
| QY | 123 | ProThrLeuValAspLysLeu---LeuValArfaspValLeuAspLys----- | 137 |
| | | :::::::::: | |
| Db | 6237 | GTACTTCAAACAGAGAGCTATTGTTAGAAAAACACAAATGAATAAACAGAAAGTTATT | 6296 |
| QY | 138 | ----CysMetGlnGluGlnLeuLeuThrIle----GluAspArgAsnArgIleAlaIleAla | 155 |
| | | :::::::::: | |
| Db | 6297 | GTCGACCTTGAAACAACACTCTCAGTGTCACAGTGAAGAGAAACCACTTCCTGACAGA | 6356 |
| QY | 156 | GluAsnAsnGlnLysGlnU----SerGlyValArfGlnLeuLeuLeuLysArgIleValGln | 173 |
| | | ::::: | |
| Db | 6357 | TTCAGTACTATGTCAAAAAACACAGGCACGTGATCACTTCTCGAAAAAATGCAAGCAG | 6416 |

| | | | | | | | | | | | | | | | | | | |
|----|------|----------------|--------------------|------------------|-----------------|----------------|----------------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| OY | 174 | Lys----- | -----Glu | Asn | Trp | Pro | Thr | Leu | Val | Leu | Pro | Thr | --- | 188 | | | | |
| Db | 6417 | AAAACACAGAGCTT | GGATGCTATCAAAAG | GAGTGTCTCTTCATTC | CGATTC | CGGTGGC | CAAG | 6476 | | | | | | | | | | |
| OY | 189 | ----- | -Gly | Asn | Ang | Leu | Val | Gln | Leu | Thr | Gly | Ser | Asp | Cys | Ser | Glu | 204 | |
| Db | 6477 | GCACAGGTGAAG | AAAAAGACGGAAC | CTCTTCAC | ACTTTTTC | ---TCTCT | GTGAGTCAG | 6533 | | | | | | | | | | |
| OY | 205 | Ser | Asn | Ala | Glu | Ile | Glu | Asn | Leu | Ser | --- | Val | Asp | Gly | Pro | Gln | 223 | |
| Db | 6534 | CTGTAA | AAAGACAAACTCAT | CTCCAGAGAAAG | CTGCAGAGTT | TCGAAATAC | CACTCAACG | 6593 | | | | | | | | | | |
| OY | 224 | Leu | Ser | Thr | --- | --- | Val | Gln | Pro | Asn | Leu | Glu | --- | 235 | | | | |
| Db | 6594 | GCACCTCTT | GCACAAAATGTGAGCTG | GAAGAAACCAATTC | GCACACTGAT | AAACAGCAAA | 6653 | | | | | | | | | | | |
| OY | 236 | Glu | Val | Trp | Gly | Met | Glu | Asn | Ser | --- | Ser | Glu | Ser | Thr | Ala | 250 | | |
| Db | 6654 | GAATTGCTT | GTCAAGGAATGTGA | AAACCTTCGACGGCC | AGAGTATG | ATATGAAAT | AGATTATGAA | 6713 | | | | | | | | | | |
| OY | 251 | Asp | Ser | Ser | Val | Val | Ser | Glu | Ser | Asp | Thr | Ser | Leu | Ala | Glu | Gly | 266 | |
| Db | 6714 | AAGCTGAAT | GTCTCCAAAGGCTT | GCAGGCCG | CACTGCTGG | AGAAAGTAC | ATTTCGCAATG | 6773 | | | | | | | | | | |
| OY | 267 | --- | Val | Ser | Cys | Leu | Asp | Glu | Ser | Leu | Gly | His | Asn | Ser | Asn | Met | Gly | 285 |
| Db | 6774 | AGGCTGAGCT | CCACACACAGAGGAAGTC | ---CAT | CACTGAGAAAG | AGCAACGCAAA | CTG | 6830 | | | | | | | | | | |
| OY | 286 | Thr | Met | Gly | Ser | Asp | Ser | Asp | Glu | --- | --- | Glu | Asn | Val | Ala | Ala | 298 | |
| Db | 6831 | AGAGTTG | CCATTGAGCCCGATG | AAAGAAAGAC | CACTGACATG | CGACAGCAAG | AACTGAAATA | 6890 | | | | | | | | | | |
| OY | 299 | Arg | Ala | Ser | Pro | Glu | Pro | Glu | Leu | Arg | Pro | Tyr | Gln | Met | Gly | Val | Ala | 318 |
| Db | 6891 | CGCGAGG | GGGAGGAATGAT | CTACCTTAAG | GAATAAAGTTG | ATGACACTTGAA | ATGGAATTGAG | 6950 | | | | | | | | | | |
| OY | 319 | Ala | Leu | Glu | Gly | Lys | Asn | Ile | Ile | Cys | Leu | Pro | Thr | Gly | Ser | Gly | Val | 338 |
| Db | 6951 | ATGTCA | ACAGAAACCGAGG | ACTAGTATCTTGAT | GTGCGAGAAATTC | CAAAACAGACATA | 7010 | | | | | | | | | | | |
| OY | 339 | Ala | Val | Tyr | Ile | Ala | Lys | --- | Asp | His | Ser | Asp | Phe | Lys | Lys | Val | Ala | 357 |
| Db | 7011 | GAGACT | AAAAACACAAATAG | AAAGAGATGGCC | AGAAAGCTG | GAAGTTT | ---GAATTAT | 7070 | | | | | | | | | | |
| OY | 358 | Lys | Val | Ile | Val | Leu | Val | Asn | Lys | Val | Leu | Val | Gln | Gly | Leu | Phe | Asp | 377 |
| Db | 7071 | CTTGCA | CGTTAAGGTGTGA | AAAAAGAAATCTG | CAACAAACAAATGCA | AAAGAAACAAAGCT | 7130 | | | | | | | | | | | |
| OY | 378 | Gln | Pro | Phe | Leu | Lys | Lys | Trp | Tyr | Arg | Val | Ile | Gly | --- | Leu | Ser | Gly | 392 |
| Db | 7131 | CAG | -----TTG | TGACAGAC | TAGACAAAGTT | ACTCTCTTCA | TTTAAAGTCT | 7184 | | | | | | | | | | |
| OY | 393 | Asp | Thr | Gln | Leu | Lys | Ile | Ser | Phe | Pro | Gln | Val | Val | Lys | Ser | Cys | Asp | 412 |
| Db | 7185 | AAGGAC | CAAGCAGAGAT | ACAGATCCAAAG | AGAACTTAA | --- | ACT | 7228 | | | | | | | | | | |
| OY | 413 | Thr | Ala | Gln | Ile | Leu | Glu | Asn | Ser | Leu | Asn | Leu | Glu | Asn | --- | --- | --- | 426 |
| Db | 7227 | GCAGTGAG | ATGCTTGAGAAATG | ATTAAAG | AGCACTAAAT | TAGGCA | GTATACCTCTTG | 7286 | | | | | | | | | | |
| OY | 427 | Gly | Glu | Asp | Arg | Lys | Val | Gln | Leu | Ser | Asp | Phe | Ser | Leu | Ile | Ile | Phe | 446 |
| Db | 7287 | GGTAC | CAACGAATATTA | TGAAGGCC | AGCAAGCAAGCTCT | CTAGCCCAAT | TAGAGCAATAG | 7348 | | | | | | | | | | |
| OY | 447 | His | Thr | Asn | Lys | Glu | Ala | Val | Tyr | Asn | Asn | Ile | Met | Ala | His | Tyr | Leu | 466 |
| Db | 7347 | CATCAGCT | GAGAAATAC | ACTT | --- | --- | GAAGAC | 7376 | | | | | | | | | | |
| OY | 467 | Lys | Asn | Asn | Arg | Leu | Lys | Lys | Glu | Asn | Lys | Pro | Val | Ile | Pro | Leu | Pro | 484 |
| Db | 7377 | AGAGCCG | CCCTAGAAAGCT | GATGAA | AAAAAGC | ACACCTGTGTGT | CTTACAAATCACTG | 7436 | | | | | | | | | | |

| | | | |
|----|------|---|------|
| Db | 7437 | AGGAGCATCATGCACATTACTTAAGGGGTAGAGTGGACAACTTGAAAGGACAATAAG | 7496 |
| OY | 485 | LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLysAlaGluGlu | 504 |
| Db | 7497 | ATAGCCGACGACAAACCAAGAGCATGCAGCTTGAGGCGACAGAAATTCACAGAGCTA | 7556 |
| OY | 505 | HisLeuLysLeu-----CysAlaSerLeuASP | 514 |
| Db | 7557 | GAGACCTTAAAGCAAAATAGAGGGATGACCCAAAGTCTGAGAGCTTGCAATTAAT | 7616 |
| OY | 515 | AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleLeuGlu | 534 |
| Db | 7617 | GTTGTACTATAAAGTCAGAAAAAGAAATCTGACAAATGAATTCAAAAAGACAAAG | 7676 |
| OY | 535 | ProCysLysLysPheAlaIleIleAlaAspAlaThrArgGluAspProPheLysGluLysLeu | 554 |
| Db | 7677 | CGAATATCTGATTAGAAATATATAATATTCATCTTTGGAAAAATATTTTGGCAAAAA | 7733 |
| OY | 555 | LeuGluIleMetThrArgIleGlnIleThrTyrCysGlnMetSerProMetSerAspPheGly | 574 |
| Db | 7734 | -----GAGCAAGAGAAAGTACAAATGAAAGAAAAATCAAGTACTGCC | 7775 |
| OY | 575 | ThrGlnProTyrGlnGlnIleThrAlaIleGlnMetGluLysLysAlaLysLysLysAsn | 594 |
| Db | 7776 | ATGGAGATCGTTCAACACAAATTAAAGAGCTCATATGAGAGAGTGGCAGCGCTGCATAT | 7833 |
| OY | 595 | ArgLysGluArgValCysAla-----GlnHisLeuArgLysTyrAsnGluAlaLeuGln | 612 |
| Db | 7836 | GACCAGAAAGCGCTTAGGCGCAAGAGCAACAAATCTAGTACAGTACAGTGTCTTGAA | 7895 |
| OY | 613 | IleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheTyrAsnGlu | 632 |
| Db | 7896 | CTT---GAGAAAGCTCATGTTGCTTCAAGCGCTTGATGAGGCCAAATATTAATATTGTT | 7955 |
| OY | 633 | GluIysAspLysLysPheAlaValIleGluAspSerAsp-----Glu | 647 |
| Db | 7953 | TTGCATCTTTCAGTAAAGGCGTCATTCAGAAAGTGAAGATGGCAAGACAAATGGAG | 8012 |
| OY | 648 | GlyGlyAspAspGluTyrCys-----AspGlyAspGluAspGluAspAspLeuLys | 664 |
| Db | 8013 | AAGAGGATGGAAGAAATCAGTAGACTGAAAAATCAATTCACACCAAGACGATTTGTC | 8072 |
| OY | 665 | LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPheGlnAsnAsn | 684 |
| Db | 8073 | TCTAACTGTCGCCAGGTCGAAAGGAGACAC-----CAACTTTGGAAGGAGCTAAAG | 8122 |
| OY | 685 | LysMetLeuLysArgLeuAlaGluAsnProGluTyr-----Glu | 697 |
| Db | 8134 | TTAGCAACTGAGAAATCTGACGCGTGAATTTGAGCAGCAGAAATCCAAATGCTACAAACCAA | 8183 |
| OY | 698 | AsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGluIleThrArgThrGluGlu | 717 |
| Db | 8184 | AATGCCCTCTTTCAGAGACACATTAGAAGTCTGACAGATTCTTACCAAGATCAATGAAT | 8244 |
| OY | 718 | SerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerLeuThr | 737 |
| Db | 8244 | GAG-----CTTGAAATTGCACAAAATGACAAATATGCTTCCTTTGTT | 8282 |
| OY | 738 | IleThrGluAsnGluLysPheAlaGluValGlyValLysAlaHisHisLeuIleThrAla | 757 |
| Db | 8283 | -----GAAAAAGTAAACAAATGACTGCCAAAGGAAACTGAGCTGCAATGCGAA | 8330 |
| OY | 758 | GlyHisSerSerGluPheLysProMetThrGlnAsnGlnIleLysGluValIleSerLys | 777 |
| Db | 8331 | ATGCATTGATGCACAGCAAAACAGCAGACCTGCACAGACAACTCATGTGACAGCAAAAT | 8390 |
| OY | 778 | PheArgThrGlyLysIleAsnLeuLeu----- | 786 |
| Db | 8391 | AGGCTAGCTGAGAGTTCCTACTTCTTGGAAGAAATAAAGACAGCAAAATTAATG | 8450 |
| OY | 787 | -----IleAlaThrValIleGluGluLysLeuAspIleLysGluLysAsn | 802 |

```

DB 8451 AAGAGCTGACACTAGAAAATAGTGAATTGAAAGAGGCTAGATTGTAACCAACAC 8510
OY 803 lIeValIleArgrTyGlyLeuValThraAngluIleAlmeValGlnAlaArgGlyArg 822
DB 8511 CAGGTGAGAAAGAGAAAGAGAAAGAGAGAGAAATAGTGAATATGATGATGATGAT 8570
OY 823 AlArGAlaAspGluSerThrTyValLeuValAlaHisSerGlySerGlyValIleGlu 842
DB 8571 GAAGCTGAAAAGAAACACAGGCTGCTTTGGACACAAACAAACAGTA:CAAGTA:AA 8630
OY 843 HisGluThrValAsnAspPheArgGlyLysMetMetTyLysAlaIleHisGlyValIle 862
DB 8631 ATCCACACA-----TACCAGAGAAATGACTTCTTAAGAA-----GAATGCTGACT 8678
OY 863 AsnMetLysProGlu-----GluTyAlaHisLysIleLeu 874
DB 8679 TCACAGAACTGAGATACACCTTTAAAGTCTAGTAAGAAAGCTCAATATCATG 8738
OY 875 GluLeuGlnMetGlnSerIleMetGlyLysMetLysThlys----- 889
DB 8739 AAGGCTACTACTAG-----ATTGGAAAGATTGAAACAAACAGATGGAATCTAATA 8795
OY 890 -----ArgAsnIleAlaLysHisTyLysAsnAsnProSerLeuIleThrPheLeuGly 907
DB 8796 TATGTAATGACTTGAAGAAAGAAATGAACGTGCCACAGGAGAAATGAATGAAATTGTTGATC 8855
OY 908 LysAsnGlySerValLeuAlaCysSerGlyLysAspIleHisValIleGlnLys----- 925
DB 8856 AATCTGTGAACAGCTGAA-----GAGGAAAGAGATACTGAG:AAAGAA:TC 8906
OY 926 -----MetHis 928
DB 8907 TCTCAACTCAAGCTGCACAGAGAAAGAAACAGGCTACTTTATGCA:ACCAAGTC 8966
OY 929 ValAsnMetThrProGluPheLysGluLeuTyrlleValAlaArgLysAsnLysAla----- 946
DB 8967 GATGAATTAACACTGAGATCAAAAGAACTGTTAAAGAAATCTTAAAGAAATCAACCAAGAG 9026
OY 947 -----LeuGlnLysLysCysAlaAspTyrlleGlnLeuAsnGlyLleIleLys 962
DB 9027 GCAGATGAATACTTGATGATAGTACTGCTCC-----TTGCTATTAAGCCATGAAGAAATAGAG 9083
OY 963 LysCysGlyGlnAlaTrpGlyThrMetMetValHis 974
DB 9084 AAAGCTAAAGAGATGTTAGACACACAAGTGGCCCAT 9119

RESULT 10
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skilman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700

```

```

; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

Alignment Scores:
Pred. No.: 3,47e-10 Length: 10136
Score: 191.50 Matches: 218
Percent Similarity: 35.60% Conservative: 185
Best Local Similarity: 19.26% Mismatches: 486
Query Match: 3.61% Indels: 243
Gaps: 45

US-09-515-363c-2 (1-1025) x PCT-US95-16216-2 (1-10136)
OY 21 ArgValLysMetTyrlleGlnValGluProValLeu----- 32
DB 5919 AAATGAAAGCATGATGAAATGGAATAATTTCTGTGATCACAGGAGTTACTCAAGAGATA 5978
OY 33 -----AspTyrlleuThrPheLeuProAlaGluValLysGluGlnIle----- 49
DB 5979 GATTTAAGTGAAGAAATGGAATATTTCTGTGATCACAGGAGTTACTCAAGAGATA 6038
OY 50 ValAlaThrSerGly-----AsnMetGlnAlaValGluLeu----- 64
DB 6039 GAACTCTGGAAGGCTCAATTCGATTAGAAATGATGATGATTAATCAATATGCA 6098
OY 65 ThrLeuGlnLysGlyValTyrlleHisLeu-----GlyTrpThrArgGluPheValGlnAla 82
DB 6099 GATATTGGAGATTAATGTGGCCAGGTGAATGACAGCTGGAAGAGAGATTTCTTAT----- 6155
OY 83 LeuArgArgThyGlySerProLeuAlaAlaArgTyrlleAsnProGluLeuAspIle 102
DB 6156 -----GTGAAATATGACCTGAGAGATC 6179
OY 103 ProSerProSerPheGluAsnAlaHisAspGlyTyrlleGlnLeuLeuAsnLysLeuGln 122
DB 6180 AGATCGAGAAAGCTAGCATTCAGCAT-----GAAGCCCTCTACTCTGAGGCTGATTTAGAG 6236
OY 123 ProThrLeuValAspLysLeu-----LeuValAlaGspValLeuAspLys----- 137
DB 6237 GTAGTCAACAGAGAACACTGTTTAAAGAAAGCAATGAAATTAACAGAGATTTAT 6296
OY 138 ---CysMetGluGluGlnLeuLeuThrIle---GluAspArgAsnArgIleAlaAla 155
DB 6297 GTCTCCTTGAAGAGAACTGCTGACAGTGTCAACAGTGAAGAAACACACTGCTGAGAGAA 6356
OY 156 GluAsnAsnGlyAsnGlu-----SerGlyValAlaGluLeuLeuLysArgIleValGln 173
DB 6357 TTAGATCTATGCTAAAAAACCAGGCACTGATGATGCTGTGAAAAATCAAGAG 6416
OY 174 Lys-----GluAsnTrpPheSerAlaPheLeuAsnValLeuArgGlnThr--- 188
DB 6417 AAACACAGAGCTGATGATCTCATCAAAAGTGTCTTCATTGCTACCTGACAGTGTGAGAG 6476
OY 189 -----GlyAsnAsnGluLeuValGlnGluLeuThrGlySerAspCysSerGln 204
DB 6477 GCAGAGGTGAAGAGAAAGAGAGAACTCTTCAAGACTTTGCTC-----TTCGATGTAGAGAG 6533
OY 205 SerAsnAlaGluIleGluAsnLeuSerGln---ValAspGlyProGlnValGlnGlnGln 223

```

```

Db 5534 CTGTTAAAGACAAACTATCTCCAGGAAAGACTGACAGTTTGGAAAAAGAGCTCATAG 6593
QY 224 LeuLeuSerThr-----ValGlnProAsnLeuGluLys 215
Db 6594 GCAGCTGCTTGCAGAAATGTGAGTGGAAACCAATTGCACAGTGAATTAAGAGAAA 6653
QY 236 GluValTrpLysMetGluAsnSer-----SerGlnSerPheAla 250
Db 6654 GAATTCGCTTCCAGGATCTGAAGCGCTGCAGCCAGACTGAGTGAATTAATTAAGAA 6713
QY 251 AspSerSerValSerGluSerAspThrSerLeuAlaGluLys----- 266
Db 6714 AAGCTGATGCTCCAGAGCGCTGGAGCGCGCACTGTGTAGAGAGAGTGGATTGCA 6773
QY 267 ----ValSerCysLeuAspGluSerLeuGluHisAsnSerAsnMetCysAspSerGly 285
Db 6774 AGCGTGGCTCAACAGAGGAGGAAATG---CATCAGCTGAGAAAGAGCAATGAGAAATG 6830
QY 286 ThrMetLysSerAspSerAspGlu-----GluAsnValAlaAla 298
Db 6831 AGAGTTCGCTTGAAGCGCTGAAAAAGACAGCTGCATCCAGAGAACTGAAAGAA 6890
QY 299 ArgAlaSerProGluProGluLeuGluLeuArgProTyrGlnMetGluValAlaGlnPro 318
Db 6891 CGCGAGCGGAGAAATGCTTCACTTAAGGATTAAGATTGAGAGACCTGGAAGAGAAATTCAG 6950
QY 319 AlaLeuGluGlyLysAsnLeuLeuLeuLeuLeuProThrGlySerGlySerThrVal 338
Db 6951 ATGTGCAAGAAACAGAGAGAGTGTGATCTGTGATCCGAGAAATTCAGAGAGAGAA 7010
QY 339 AlaValTyrLeuAlaLys---AspHisLeuAspLysLysLysLysLysLysLysLysLys 357
Db 7011 GAACACTTAAAAACAAATAGAAAGAGATGCCAGAGCTGAAAGATTTCGATTAGAC 7070
QY 358 LysValIleValLeuValAsnLysValLeuLeuValGluLeuPheArgGluPhe 377
Db 7071 CTGTGACGTTAAGTCTGAAAAAGAAATCTGCACAAATTCACAAAGAAACAAAGGT 7130
QY 378 GlnProPheLeuLysLysTrpTyrArgValIleGly-----LeuSerGly 392
Db 7131 CAG-----TTGTGACAGAACTAGACAGTTACTCTCTTCATTAAAGTCTGTAAGAA 7184
QY 393 AspThrGlnLeuLysLysSerPheProGluValValLysSerCysAspLeuLeuSer 412
Db 7185 AAGGACCAACAGACATACAGATCAAAAGAGAAATCTTAA-----ACT 7226
QY 413 ThrAlaGlnIleLeuGluAsnSerLeuAsnLeuGluAsn----- 426
Db 7227 GCAGTGGAGATGCTTCAGATTCAGTTAAAGAGCTTAATAGCGAGTAAAGCGCTTGT 7286
QY 427 GlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuLeuLeuLeuLeuLeuLeuLeu 446
Db 7287 GGTGACCAAGAAATTAATGAGGCCACAGAAAGAGTCTAACCACCAATAGAGAGAGAG 7346
QY 447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLys 466
Db 7347 CATCAGCTGAGAAATAGCAAT-----GAAAAATG 7376
QY 467 LysAsnAsnArgLeuLysLysGluAsnLysProValIleProLeuProGluIle----- 484
Db 7377 AGAGCCCGCTAGAGCTGATGAAGAAAGACAGCTGTGTCTTACAAACAGCTGAAGAA 7436
QY 484 ----- 484
Db 7437 AGTGAGCATGAGAGATTAATTACTTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7496
QY 485 LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLysAlaGluGlu 504
Db 7497 ATACCGAGGACAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7556
QY 505 HisIleLeuLysLeu-----CysAlaAsnLeuAsp 514

```

```

Db 7557 GAGACCCATAAAGCAAAATATACAGGATGACCCAAAGCTGTGAGAGTGTGAAATAGAT 7616
QY 515 AlaPheThrIleLysThrValLysGluLeuAspGlnLeuLysAsnGlnIleGlnGlu 534
Db 7617 GTTGTACTATTAAGGTCAGCAAAAAGAAATCTGCACAAATGAAATTAACAAAAGAGAG 7676
QY 535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGluLysLeu 554
Db 7677 CGAATATCTGAATTAAGAAATTAATTAATTCATATTGAAAATATTTTGCAGAGAAATA 7733
QY 555 LeuGluIleMetThrArgIleGlnThrTyrCysGlnMetSerProMetSerAspPheGly 574
Db 7734 -----GAGCAAGAAAGTACAGATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 7775
QY 575 ThrGlnProTyrGluGlnIleProAlaIleGlnMetGluLysLysAlaLysLysLys 594
Db 7776 ATGGAGATGCTTCCAAACAAATTAAGAGCGCTCAATGAGAGAGTGGAGCGCTGTATAT 7835
QY 595 ArgLysGluArgValCysAla-----GluHisLeuArgLysTyrAsnGluAlaGln 612
Db 7836 GACCAAGAGCGCTTAAGGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7895
QY 613 IleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheTyrAsnGlu 632
Db 7896 CTT---GAGAGAGGCTCAGTGTCTACAGAGCGCTTGATGAGCGCAAAATATATATATGTT 7952
QY 633 GluLysAspLysLysPheAlaValIleGluAspAspSerAsp-----Glu 647
Db 7953 TTGCAATCTTCAAGTGAAGAGCGCTCATTCAGAAAGTGAAGAGAGAGAGAGAGAGAG 8012
QY 648 GlyGlyLysAspGluTyrCys-----AspGlyAspGluAspLysPheLysLys 664
Db 8013 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8072
QY 665 LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhePheGluAsnAsn 684
Db 8073 TCTAAAGCTCCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8123
QY 685 LysMetLeuLysArgLeuAlaGluAsnProGluTyr-----Glu 697
Db 8124 TTAGAACTAGAAATCTGACGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8183
QY 698 AsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgTrpGlu 717
Db 8184 AATGCTCTTTCAGAGACACATTAAGAGCTGCAGAGCTTCTTACAGAAATTAAGAAAT 8243
QY 718 SerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSer 737
Db 8244 GAG-----CTGAAATTCAGCAAAATGAGCAAAATGCTCTTGT----- 8282
QY 738 IleThrGluAsnGluLysPheAlaGluValGlyValLysAlaHisLeuLeuAla 757
Db 8283 -----CAAAAGATTAAGAAATGAGCTGCAAAAGAAAGAGAGAGAGAGAGAGAGAG 8330
QY 758 GlyHisSerSerLysPheLysProMetThrGlnAsnGlnGluLeuValIleLys 777
Db 8331 ATGCATGAGATGCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8390
QY 778 PheArgThrGlyLysIleAsnLeu----- 786
Db 8391 AGCGTAGCTGAGAGAGTGTGAGTGTGGAAGAAATTAAGAGAGAGAGAGAGAGAGAGAG 8450
QY 787 -----IleAlaThrThrValAlaGluGluGlyLeuAspIleLysGluLysAsn 802
Db 8451 AAGGAGCTACACTAGAAATATAGTGAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8510
QY 803 IleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArg 822
Db 8511 CAGGTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8570
QY 823 AlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerCysLysGlyValIleGlu 842
Db 8571 GAAGGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8630

```

QY 843 HIGLUThrValAsnAspPheArgGluLysMetMetLeuValAlaLeuHisValGln 862
 Db 8631 ATCCAGACA-----TACGAGAGAAATTGACTTCTTAAGAA---GAATGCTCAGT 8678
 QY 863 AsnMetLysProGlu-----GluValAlaHisLys(I)Leu 874
 Db 8679 TCACAGAGAGCTGAGATAGACCTTTAAAGTCTAGTAAAGACAGCTCAATATTCATTC 8738
 QY 875 GluLeuGlnMetGlnSerIleMetGluLysLysMetLysThrLys----- 889
 Db 8739 AAAGCTACTACTCAG---ATTGGAAAGAAATTGAACAAACCAAGATGACAAATCTAANA 8795
 QY 890 -----ArgAsnIleAlaLysHisLysThrLysAsnAsnProSerLeuLeuPheLeuGly 907
 Db 8796 TATGTAATTCAGTTGAAGAAAGAAATGAACGTGCCAGAGGAAATGAAGTTGTGATC 8855
 QY 908 LysAsnGlySerValLeuAlaLysSerGlyLysAspIleHisValIleGluLys----- 925
 Db 8856 AATTCCTGTAACAGCTGGAA-----GAGGAAAGAGAGATCTGCAGAAAGAACCTC 8906
 QY 926 -----MetHisHis 928
 Db 8907 TCTCACTTCAAGCTGCAGACAGAGAAAGCAAGAAACAGTACTGTATGAA ACCAAGGTC 8966
 QY 929 ValAsnMetThrProGluPheLysGluLeuTyrIleValArgGluAsnLysAla----- 946
 Db 8967 GATGAATTAACAACTGAGATCAAGAAAGAAAGAAACTCTTCAAGAAAGAACCAAGAG 9026
 QY 947 -----LeuGlnLysLysCysAlaAspTyrGlnIleAsnGluGluIleLeuGly 942
 Db 9027 GCAGATGAATCTGATGATAGTACTCTCTCC---TTCCTTATTAAGCCATGAAGTTAAG 9083
 QY 963 LysCysGlyGlnAlaIleTyrGlyThrMetMetValHis 974
 Db 9084 AAAGCTAAAGAGATGTAGAGACACAAAGTGGCCCAT 9119

RESULT 11
 US-08-685-576-5
 Sequence 5, Application US/08685576
 Patent No. 5906819
 GENERAL INFORMATION:
 APPLICANT: Kalibuchi, Kozo
 APPLICANT: Iwamatsu, Akihito
 APPLICANT: Nakano, Takeshi
 APPLICANT: Ito, Masaaki
 APPLICANT: Takahashi, No. 5906819uak1
 TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.10
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/685,576
 FILING DATE: 24-JUL-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 7-325129
 FILING DATE: 20-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-17150
 FILING DATE: 05-JAN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-131206

FILING DATE: 26-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16887/843
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4363 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..4164
 US-08-685-576-5

Alignment Scores:
 Pred. No.: 1.71e-10
 Score: 188.50
 Percent Similarity: 35.43%
 Best Local Similarity: 18.51%
 Query Match: 3.55%
 DB: 2
 Caps: 59

US-09-515-363c-2 (1-1025) x US-08-685-576-5 (1-4363)

QY 71 TRPHISLeuGlyTRPThrArgGluPheValGluAlaLeuArgArgThrGlySerProLeu 90
 Db 1075 TGGCAT-----TGG-----GATTAACATTAAGAAAGGACGACCTCTCTCA 1113
 QY 91 AlaAlaArgTyrMetAsnProGluLeu---ThrAspLeuProSerProSerThrIleuAsn 109
 Db 1114 GTA-----CCTGAAGCTCAGCACTGACATAGACGAGCAATTTTATGAC 1158
 QY 110 AlaHisAsp-----GluTyrLeu---GlnLeuLeuAsnLeuLeu----- 121
 Db 1159 ATTGAAGATGACAAAGAGATGTAGAAACCTTCCCAATTCCTAAAGCTTTTCTT 1218
 QY 122 -----GlnProThrLeuValAspLysLeuLeuValArgAspValLe 135
 Db 1219 CAGCTGCTTCATCGATGATTTACTACTATAGA-GAAATTTATATTATAGTGA---TC 1274
 QY 135 AsnLysCysMetGluGluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAl 155
 Db 1275 TCCATCTTGTAGAGAAATGAT-----TCCATTAACATCAAGAAA----- 1314
 QY 155 AsnLysAsnGlyAsnGluSerGlyValArgGluLeuLeuLysArgIleValGlyG 175
 Db 1315 -----AATGAAGAAAGT---CAACAGATTCAGAAAAAAGCTGTATAATTAGA 1358
 QY 175 AsnTrpPheSerAlaPheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValG 195
 Db 1359 AGAACATCTTAC-----AATGAGATGCAAGCCAA 1388
 QY 195 GlnLeuThrGlySerAspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnVa 215
 Db 1389 ACAGGAAGTGAAGCAAGCAATCTGTTAATCTGCGCTAGCAAAAAAGCAAG-AAAG- 1446
 QY 215 AspGlyProGlnValGluGluLeuLeuSerThrThrValGlnProAsnGluGly 235
 Db 1447 -----GAGCTTAGAGAGAGAGATT-----ACCTTAGCAAGAAAGTGTAATC 1487
 QY 235 sGluValTrpGlyMetGluAsnAsnSerSerGluSerSerPheAlaAspSerSerValVa 255
 Db 1488 ACCATTAGACAGCTTAGAAAGAAAGAAAGGCGCTTCTTCAGCAGCAAAAATGCGAATATCA 1547

QY 255 lsergluserAspThrSerLeu-----AluGluclySerValSerCysLeu 270
 Db 1548 GAGGAAAGCTGATCATGAGACAGACAAAGAAATTTGAAATATGATCTTACAGCTT 1607
 QY 270 uaspgluserLeu-----GlyHisAsnSerAspMetCysSerAspSe 284
 Db 1608 AAAAGATCAACTGAGATTTGAAAAAAGAAATCAAAATCTCAAAATATCACTGAGAA 1667
 QY 284 r---GlyThrMetGlySerAspSerAspGluGluValAlaAlaArgAlaSerProG 303
 Db 1668 AGTGAATCAACTGCAGACAGCAACACTGCATGCAACCAATGCTTACGCGCAAGAGCTCTCA 1727
 QY 303 uproGluLeuGlnIleuArgProtyrGlnMetGluValAlaGlnProAla----- 319
 Db 1728 TACTGACGCCGCTTAAGCAAAAACCCAGCAGAAAGTTCAAAACATTTACAGAGCTGCA 1787
 QY 320 -----LeuGluGlyLysAsnIleIleIleCysLeuProThrGlySerG 334
 Db 1788 ATCTAACAAATAGAGATCTCAAGATTAATAAACTGCTG-----CTGAGACATGCCAAGTT 1841
 QY 334 ylystrArgValAlaValTyrIleAlaLysAspHisLysAspLysLysLysLase 354
 Db 1842 AAAACTTGAAAAAGCAATTTATCAATCTTCACTCAGTCACTCTAGAAATCTGAAGAGGAGATCG 1901
 QY 354 rgluproGlyLysValIleValIleuValAsnLysValIleuValGluIleuPheAr 374
 Db 1902 AACCCATGATCGATCGAGATATATT----- 1923
 QY 374 glyslupheGlnProPheLeuLysLysTyrPyrArgValIleGlyLeuSerGlyAspTh 394
 Db 1924 -AATGATTTACAAAGT-----AGAACATGCGCTTAAGCAAGAT-- 1962
 QY 394 rgluIleuLysIleSerPheProGluValValLysSerCysAspIleIleIleSerThrAl 414
 Db 1963 -----TTAAAGAACGGTAAATCTTTAAGCGAAGT 1994
 QY 414 aglnIle-----LeuGluAsnSerLeuLeuAsnLeuGluAsnGlyLysuSpAl 430
 Db 1995 AGAAGTGGAGAACAGAACAACTTCAAGAGATTTACTGATTTGAAAAG--GAAAAGAG 2051
 QY 430 aglyValGlnLeuSer---AspPheSerLeuIleIleIle-----As 443
 Db 2052 CAACATGGAAATAGATATGATGACATACCAACTAAAGTTATTCACGACAGCTTAGACACAGA 2111
 QY 443 pglucySHisHis-----ThrAsnLysGluAlaValLysLysAsn 456
 Db 2112 AGAAGCTGAACATTAAGCCCAAAAGCAGACACTAGACAGATAAAATATAGCACTATGCTC 2171
 QY 456 nileMetArgHisTyrIleuMetGlnLysLeuLysAsnArgLeuLysLysGluAsnLys 476
 Db 2172 CATC-----GAAAGAGCCAAATCATCGAAGCCATGAAATATGCGAGAA 2213
 QY 476 sproValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyGlyAl 496
 Db 2214 GAAGCTCTTCGAG-----GAAAG 2231
 QY 496 arthrGlyAlaLysAlaGluGluHisIleLeu-----LysLeuLysAlaSerLeu 513
 Db 2232 AACTTTAAACAGAAAGTGAAGAACTATTGCTAGAAAGCTGAGAAAGATTTCTCTCATTT 2291
 QY 513 uasPalaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysHisGlnIleG 533
 Db 2292 AGACTGT-----GACCTCAACAGTCACAGCAGACAGAAATATAA 2327
 QY 533 ngluproCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProLysLysGlyLys 553
 Db 2328 TGAGCTCTTAAACAG-----AAAGATGCTCAATATAGAGAT--CTAGAAACCT 2375
 QY 553 sleuLeuGluIleMetThrArgTyrGlnIleThrTyrCysGlnMetSerProLysSerPhe 573
 Db 2376 GACATTTAAATAGAGCAAGAACTGAGAACCGCTC-----CTTACATTAATGAGCTT 2429
 QY 573 eglYThrGlnProtyrGluGlnIleProAlaIleGlnMetGluLysLysAlaLysLysG 593

Db 2430 GAAGATGCAAAACACACAGGTTAACACACTAAATAATGTCAAGAAAAACAGTTAAACAGAGA 2489
 QY 593 yasnArgLysGluArgValCysAlaGlnHisLeuArgLysTyrAsnGluAlaLeu----- 611
 Db 2490 AAATTAACATCTCATGAGAAATGAAAAATGAACTTGGAAAAACAAATGCTGATCTTGAAA 2549
 QY 612 -----GlnIleAsnAspThrIleArgMetIleAspAlaIleTyrTh 624
 Db 2550 AGAAGCTCAGATGCAGATGGGCAAAATGAAAGAGCTCCAGAGATGCTGCAAGCAGACAGA 2609
 QY 624 rHisLeuGlnThrPheTyrAsn-----GluGlyLys 634
 Db 2610 GTATTCTCAACCTTTATATAAACCAAGTTAGGAGCTTAAGAAAGATGTGAAGAAAAA 2669
 QY 634 sasLys-----LysPheAlaValIleGluAspAspSerAsp-- 646
 Db 2670 GACCAAACTTGSTTAAGATTCAGCAGCAAGAAACAGAAATTTACAGATGAACGGAGCTC 2729
 QY 647 -----G 647
 Db 2730 TTGGCTGCCCACTGAGATCCTTGACCAAAACAGATTTGTGACAACTGGCTTTGTC 2789
 QY 647 uclGlyLysAspArgLysTyrCysAspGlyAspGluAsp-----GluAspAspPro 663
 Db 2790 AATGCTGAAGAAACAATATTTGATTTGAAAAAGAAAGATCATGAAGAGCTCAGAT 2849
 QY 663 ulysLysProLeu-----LysLeuAspGlnThrAspArgPheLeuTyrTh 678
 Db 2850 CAAGAGATGATGGCTAGACACAAACAGAACTTACGCAAAAAGATGTACAAATTTGTC 2909
 QY 678 rleuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluTyrLys 698
 Db 2910 TCTT---GAGGAACCTAATAGCAGCACTAAGTATGATGTGCTGCAATCTTCCAAATGAGAA 2966
 QY 698 ngluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgThrGluGlu-- 717
 Db 2967 AGAAGATTAATTAACAAATTAAGAAATGATGTCAAGCAAGCACTGCAAGATTTGAAGATGA 3026
 QY 718 -----SerAlaArgGlyIleIlePheThrArgThrArgLysAsnAlaValLase 735
 Db 3027 AGAATTAACCGGACGACCTATT-----MAAGCACTTTAGCA 3065
 QY 735 rgluThrIleThrGluAsn-----GluLysPheAlaGluValG 748
 Db 3066 GCAGCTATTACAGAAAGAAACACTCAAACTCAAGCTGATGATGATGATGATGATGATGATG 3123
 QY 748 yValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGluPheLysProMetIleG 768
 Db 3124 -ATGATCGAAAAAGAACTGTCAAGCGTGTGATATGACAGATGTGCGG--AGAAAGCA 3179
 QY 768 nasngluGlnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeu----- 786
 Db 3180 GAAGGAAATAGAAAGCTCATATGAGCTTAAATGTCAAGCTGAGAAATATGACAGCA 3239
 QY 787 -----IleAlaThrThrValAlaGluGlyLys 796
 Db 3240 GATGATCAAGTATCAGAAAGAACTGAAATGCAAGGAGCAAAATAGCTGAGAG-- 3294
 QY 796 uasPrlLysGluCysAsnIleValIleArgTyrGlyLeuValThrAsnGlnIleLase 816
 Db 3295 -----AGCAAGATTCGAATTAAGTCAAGATGATGATGATGATGATGATGATGATGAT 3347
 QY 816 tValGlnAlaIleArgValArgAlaVal-----AspGluSerThrTyrTyr 831
 Db 3348 TGAGAGCTGCGGTCAACACTCCAAAGCTTGCATATGCTGCTGAGTGTAGTGTAGTGTAGT 3405
 QY 831 lleuValAlaHisSerGlySerGlyValIleGlnHisGluThrValAsnAspThrArg 851
 Db 3406 -----GGCAGTGGACAGGAGATGCTGAGAGAGATGATGATGATGATGATGATGATGAT 3449
 QY 851 ulysMetMet-----TyrLysAlaIleHisCysValGlnAsnMetLysProGlnTyr 869

```

Db 3450 ATCAAGATTAGAGATGCTTTCATTCCTGCTAGCAAAACACCTAAGCAATTCGATG 3509
OY 869 TAlHlHlLylslleuGluleuGlInMetGlnSerlIleMeGlulysLysLeuLysThrLy 889
Db 3510 GGTAAAAAGATGTGATTGATTAGCAGCTAAGAAAGATCTTCTTATTCACAGCTGACAGCA 3569
OY 889 sArGAsnllleAlalysHlstrYlYsAsnAsnProSerleuIleThrPhoLysLysAs 909
Db 3570 TAAAGAA-----CAATCCAACTCTTACATGTTTATGATTTAGACAAATTT 3614
OY 909 nCySeSerValleuAlaCySerSerGlyGluAspIleHisValIleGlulysMetHisIsla 929
Db 3615 ATTTCATGTCGACAGCACTTACACAGACAGATGTCAT----- 3651
OY 929 lAsnMetThrProGluPhelLysGluLeu-----TyrIleValArgGluAsnLy 945
Db 3652 -----AGAGCAGATCTCTAAGAAATTCAGAGATATTCAGATTCGTATGCCAATGA 3704
OY 945 sAlaLeuGlulLysLysCysAlaAspTyrGlnIleAsn-----GlyGlu----- 3760
Db 3705 AGGAGAAAGTAAAGAGACAAAGAAATTCAGAGTGGAGCCGTTGGAGAAATCTAATTA 3764
OY 960 eIlleCyLysCyGlyGlnAlaIleTyrGlyThrMetMetValHis----- 3774
Db 3765 TATTGCCACAGAGCAGATGAGTTTATTCCTACTCTTATCATTTCCCAACCAACTGGA 3824
OY 974 ----- 3824
Db 3825 GCGTTGATAGAGCCCTGTGGCAGCATGTTAAGCTCTCTCTGCTTTGAGTGGCCGCG 3884
OY 975 -----LysGlyLeuAspLeuProGly 981
Db 3885 TTGGCATATTAAAGTGCATAAAGATCATATGACAAAGAGAGACATTTAGACACTTG 3944
OY 981 gLeuLysIleArgAsnPhenValValPheLysAsn-----AsnSerThr 996
Db 3945 C---AAAGTATTATTCATATTTTCACGCGCAAGAAATCTGTACTACTAAATTCAG 4001
OY 996 tLysLysGlnTyrLysLysTyrVal 1004
Db 4002 AGAAGACAG---CAGAGTGGCTT 4023

RESULT 12
US-09-643-597-117
: Sequence 117, Application US/09643597
: Patent No. 6426072
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Fan, Liqun
: APPLICANT: Kalos, Michael D.
: APPLICANT: Banquer, Chaitanya S.
: APPLICANT: Hosken, Nancy
: APPLICANT: Fanger, Gary R.
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Henderson, Robert A.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF INHERITED AND ACQUIRED LUNG CANCER
: FILE REFERENCE: 210121.455C11
: CURRENT APPLICATION NUMBER: US/09/643,597
: NUMBER OF SEQ ID NOS: 369
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 117
: LENGTH: 6921
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-643-597-117

Alignment Scores: 5.05e-10 Length: 6921

```

```

Score: 187.50 Matches: 214
Percent Similarity: 34.688 Conservative: 164
Best Local Similarity: 19.638 Mismatches: 375
Query Match: 3.538 Indels: 337
DB: 4 Gaps: 51

US-09-515-363c-2 (1-1025) x US-09-643-597-117 (1-6921)
OY 41 GluValLysGluGlnIleGlnArgThrValAlaIleSerGlyAsnMetGlnValGlu 60
Db 1313 GAAATACAGCGCTGAAATAGTAAACCTAGCCAGCAGCACTTGATCAACAG-----AAG 1366
OY 61 LeuLeuLeuSerThrLeuGlnLysGlyValIlePheLysGlyTyrThrArgGlnIleVal 80
Db 1367 ATGCTGTGTCGCAATATGAAATGAAACAGACCAAAATGACAGAGTGTCAAAATATGCA 1426
OY 81 GluAlaLeuArgArgThr----- 86
Db 1427 GAACAGTACTACAGCTACAGTGAAGCACTATGAAATTAACAACATGACCTACGACACATG 1486
OY 87 -----GlySerProLeuAlaAlaArgTyrMetAsnProGluValThrAsp 101
Db 1487 GTAGATTCAACAAACAAATCTCAGTGAAGACCGCAAGATGACAGATTCA--GCAAT 1543
OY 102 LeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuAsnLeuLeu 121
Db 1544 CTCATTATTCAAGAGCTTCAGTGAAGCACTGACCTGATCTGCTGCTCAATTCATG 1603
OY 122 GlnProThrLeuValAspLysLeuValArgAspValLeuAspLysCysMetGlnGlu 141
Db 1604 ACACATATATAT-----AAATTGCTGTGATTCATTAAACAGCTGCAAGAGAG 1654
OY 142 GluLeuLeuThrIleGluAsp-----Arg 149
Db 1655 GAGATTAAAGGTGATGAGAGAGACTCTGAAACATGGGGCATATTCAGATCTGCTTACGGT 1714
OY 150 AsnArgIleAlaAlaAlaLeuLysAsn-----GlyAsnGluSerGlyValIleGlu 166
Db 1715 CAGAAAGCAACAGCTGCTGATGATGACAACTTACAGAGAAACATAGTAGTGTGAAAGA 1774
OY 167 LeuLeuLysArgIleValGlnLysGluAsnTyrPheSerAlaPheLeuAsnValIleVal 186
Db 1775 ATGCTAGCTGAACTAAGAAACAAAGTCCGAGTAGAGAAAGAACTCCGAAAGTCCAG 1834
OY 187 GlnThrGlyAsnAsnGluLeu-----ValGlnGluLeuThr----- 198
Db 1835 GAGCTGCAGAAATGAAATGATGAGAAAGCAGAGAAATAGAGATATCTCTCTAG 1894
OY 199 -----GlySerAspCysSerGluSerAsnAlaGluIleGluAsnLeuSer 213
Db 1895 AAGATTAAAGGCTGAAAGTGAAGCCAAAGCAAGTACCCAGAGGAACTTGAAACCTTGTAGA 1954
OY 214 GlnValAspGlyProGlnValGlu-----GluGlnLeuLeuSerThrThrValIleThrAsn 232
Db 1955 GAGAGAGAAAGCGCTGAAAGAAAGAACTGAGCGCGTAGAGCAAGCTCACCATAGATCTGAG 2014
OY 233 -----LeuGlnLysGluValIleTyrGlyMetGluAsnAsnSerSerGluSer 247
Db 2015 GCTAAAGAGCTGCGGTGAGAGAAAGCACTCTGAAATTTCCAACTGAGTGAAGCAAC 2074
OY 248 SerPheAlaAspSerSerVal-----ValSerGluSerAspThrSerIleValIle 264
Db 2075 ACCTTTACAGAGCAACACTGAGATCATCTTAAACAAAGATTTAAGTCT----- 2128
OY 265 GlySerValSerCysLeuAspGluSerLeuGlnHisAsnSerAsnMetGlySerAspSer 284
Db 2129 -----AATGATTTGGAGCAACAAACAAATTAATTAATTAATTAATTAATTAAT 2167
OY 285 GlyThrMetGlySerAspSerAspGluLysValAlaAlaArgAlaSerThrAsnPro 304
Db 2168 TTAAGAGAAAGAGAGCAATGAGAGAAAGCACTGTGAGATTAAGCAGAGATTAAGAAA 2227
OY 305 GluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGlnLysGlyAsn 324

```


OY 921 HisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLeuTyrIle 940
 Db 4022 CTGTCTCTAGAAAGACACTGGACACTTACCCAGACAGCCAGATCTCTCTGTGAGATGG 4081
 OY 941 ValArgGluAsnLysAlaLeuGlnLysLys 950
 Db 4082 ACTCAAGAACACACAGCATTGGAGAGAGAG 4111
 RESULT 13
 US-08-961-527-142
 ; Sequence 142, Application US/08961527
 ; Patent No. 6420135
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,527
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8512
 ; TELEFAX: (301) 309-8504
 ; INFORMATION FOR SEQ. ID NO: 142:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5020 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-961-527-142
 Alignment Scores:
 Pred. No.: 3,26e-10 Length: 5020
 Score: 187.00 Matches: 157
 Percent Similarity: 34.53% Conservative: 132
 Best Local Similarity: 18.76% Mismatches: 263
 Query Match: 3,528 Indels: 285
 Db: 4 Gaps: 39
 US-09-515-363c-2 (1-1025) x US-08-961-527-142 (1-5020)
 OY 309 ArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGly-----LysAsnIleIle 326
 Db 1097 CAAGCCCAAGCTATGAGACAGTTGGTGATTAACATTGAGGGGAGCAAAACCTCAGATT 1156
 OY 327 IleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrIleAlaLysAspHis 346
 Db 1157 CTGATGGGGGGGAGCTGGAACAGGAGAGACC-----TATCTATGAGTCAAGTC 1204
 OY 347 LeuAspLysLysLysLysAlaSerGluProGlyLysValIleValLeuValAsnLysVal 366
 Db 1205 ATTCTAAAGTCAATAAACCAACT-----CTGTATTGTCTCAACATATA 1249

OY 367 LeuLeuValGluGlnLeuPheArg-----LysGluPheGlnProPheLeuLys rStrp 384
 Db 1250 ACTTGCTGTGCTGAGCTCTATGGGAGATTAAAGAAATTTTCCCT----- 1294
 OY 385 TyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe----- 400
 Db 1295 -----GAAATGCAGTTGAGATTTTCTGATCTTACTATCATTAATAC 1336
 OY 401 ---ProGluVal---ValLysSerCysAspIleIleIleSerThrAlaGlnIleProGlu 418
 Db 1337 CAGCCAGAGCGCTATGCTCCCTCTAGCCGATACCTATATGACAGAGATAGTCTTCATAT 1396
 OY 419 AsnSerLeuLeuAsnLeu---GluAsnGlyGluAspAlaGlyValGlnLeuSerAspPhe 437
 Db 1397 GACGAGATTGACAAAGCTTGCCGACCTACCTACGCTTGGAGCTGATATATGTT 1456
 OY 438 SerLeuIleIleIleAspIleCysHisHis-----ThrAsnLysGluValValTyr 454
 Db 1457 ATTGTGCTGGCTCAGTCTTGTATCTATGCTTTGGGTTGCCCAAGAAATACCTGAT 1516
 OY 455 Asn-----AsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 Db 1517 AGTGTGCTAGTCTGCTGCTGCTGCTAGAGATTCTCGATGAATACCTTGAATGACTTG 1576
 OY 467 -----LysAsnAsn-----ArgLysLys 472
 Db 1577 GTGATATTACCTATTGAACGTTATATGATTATTCACAGCGGAGAAATTGCGGCTTGT 1636
 OY 473 LysGluAsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGly 492
 Db 1637 GGGGATGTGTAGATTTTCCA----- 1660
 OY 493 ValGlyGlyAlaThrLysGlnAlaLysAlaGlnGlnHisIleLeuLysLeuCysAlaAsn 512
 Db 1661 -----GCTCCGAGATGAACAT----- 1678
 OY 513 LeuAspAlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLys----- 529
 Db 1679 -----GCCTTTCAGATGAATATTTTGGAGAGCAAAATGCGATTGCTGAACATTCAG 1732
 OY 530 -----AsnGlnIleGlnLeuProCysLysLysPheAlaIleAlaAspAla----- 544
 Db 1733 GCTGTGACAGGTCAGGTGTGGGAGAGTGGATCTTTAGCCATTTTCCCAAGTGAACAC 1792
 OY 545 -----ThrArgGluAspProPheLysGluLysLeuLeuGlnIleMetThrArgIleGln 562
 Db 1793 TTGTGACCAATGACGACACATGAGATTGCGCATTTGCAAAATGTCAGCCGAGTTGCA 1852
 OY 563 ThrTyrCysGlnMetSerProMetSerPheGlyThrGlnProTyrGluGlnIleAla 582
 Db 1853 -----GAAATATTAAGT 1864
 OY 583 Ile---GlnMetGluLysLysAlaAlaLysLysGlyAsnArgLysGluArgVal----- 599
 Db 1865 GTCTTTGAAAGAGAGGTAAGTCTGTTGAAGCCGACCTTTGAAACAGCGACATATAT 1924
 OY 600 CysAlaGlnHisLeuArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleLeuMet 619
 Db 1925 GATATGCAAAATGTCGCGATGATGGCTATACCAATGAGGTTGAAAAATATATTCCTG 1984
 OY 620 IleAspAla-----TyrThrHisLeuGluThrPheLysAsnGln 632
 Db 1985 ATGATGAGCAGCAGGAGGAGAGAGAGCTCTTATACGCTTTCGACTTCTTCATATGAT 2044
 OY 633 GluLysAspLysLysPheAlaValIleGluAspAspSerAspGlu-----GlyLysP 650
 Db 2045 -----TTCTGATTATGATTGACGAGATCATATGACATATGAGAAATC 2089
 OY 651 AspGluTyrCysAspGlyAspGluAspGluAspLys----- 663
 Db 2090 AAGGCGATGTAACATGAGACCGTCCGGTAAAGAAATGCGTTAATTATAGTTTCCT 2149
 OY 664 -----LysLysProLeuLysLeuAspGluThrAspArgIleLeuMet 677

```

Db 2150 TTGCGCTGCTTGGACAATGCTCTCCGCGGAGAGATGAGAT GACGTTAT 2209
Oy 678 ThrPhePheGluAsnLysMetLeuLysArgLeuAlaGluAsnProGluTyrGlu 697
Db 2210 CAGATGTTTACCTTCA-----GCCGACCTGTTGACTATGCA 2248
Oy 698 AsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTyr----- 712
Db 2249 AATGACAG-----ACCGAGACAGTATGACCAATACTCTCCCAACGCA 2296
Oy 712 ----- 712
Db 2297 CTCTGATCCAGAGGTGAGAGTCCGCTCCGACTATGAGACAGATGATGATCTTGGAT 2356
Oy 713 -----ThrArgThrGluGluSerAlaArgGlyIleIlePheThrLysThrArgGln 729
Db 2357 GAAATCAATGCCCGCTTAAAAAATGAGCGTACCTTATCAACCTTGACCAAGAA 2416
Oy 730 SerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLysPheAlaGluVal 749
Db 2417 ATGGCAGAGATTGACCCAGTAC-----TTCAAGGAATAGGCTATC 2458
Oy 750 LysAlaHisHisLeuIleGlyAlaGlnHisSerSerGluPheLysProMetThrLysAsn 769
Db 2459 AAGCTCAAGTACATG-----CACTCGATATCAAGACCTTGG----- 2494
Oy 770 GluGlnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeuValIleAlaThr 789
Db 2495 GAACGGAGCGAGTTATCCGTCGTCGCTGACCTGGTGTGTTCATGATCTTGTGGAAAT 2554
Oy 790 ThrValAlaGluGluLysLeuAspIleLysGluCysAsnIleValIle----- 805
Db 2555 AACCTGCTCCGTGAAGAAATGACCTCTCGAAGTGAAGCTGCTACTATCTCGATGCT 2614
Oy 806 ---ArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArg 824
Db 2615 GACAGAGAGAGTTCTCTGCGACGAACGTCGCTATCCAGACACATGAGCTGCTGCA 2674
Oy 825 AlaAspGluSerThrTyrValLeu-----ValAlaHisSerGlySerGly 839
Db 2675 CGTAATAGCGCAAGTATGATCATGATATGCGGACACGCTTAACAGTATATGCAAGT 2734
Oy 840 ValIleGluHisGluThrValAlaAsnAspPheArgGluLysMetLeuTyrLysAlaIleHis 859
Db 2735 GCTATGCTAT-----GAACTGCCCGCCCGCGCAAAATCCAGATGCTATTAAGAAAGCAT 2791
Oy 860 CysValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuGlnMet-- 878
Db 2792 -----CGTATCGTTCACAAACCATCAAGAAAGAAATCCGCTACTTATTGCTGATG 2842
Oy 879 -----GlnSerIleMetGluLysMetLysThrLysAspAsnIleAlaLysHisTyr 897
Db 2843 ACCAAGCGACCTGCTAAGGAGAAACAAAGAGTGCATCAATGACCCCAACAAACA 2902
Oy 897 AsnAsnProSerLeuIleThrPheLeuLysLysAsnGlySerValLeuLysSerGln 917
Db 2903 GAGC-----GCAAGGAACCTACTCA----- 2921
Oy 917 yGluAspIleHisValIleGluLysMetHisValAsnMetThrProIlePheLysGln 937
Db 2921 ----- 2921
Oy 937 uLeuTyrIleValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspLysIle----- 954
Db 2922 -----AAAGCTTGAGAAACAAACAAATGCAAGAGAGTGTAGTG 2959
Oy 955 -----GlnIleAsnGluGluIleLysCysLysCysGlyValAlaIlePro-- 968
Db 2960 CTGACTTTGAAGTACAGAGCTGATGATATGATGCTGAAGTCAAGAGCTTGGAT 3019
Oy 969 -----G 969

```

```

Db 3020 TAGGGGAATAGTATGATTTATTTAAGAAAGTTAAGAAAGATTGATGTCATATG 3079
Oy 969 yThrMetValHis-----LysGlyLeuAspLeuProCysLeuLys 983
Db 3080 GAAATGGCTTATTCACACCTTATTCACAGTTTGGAAGAAAGTATGATGCTCCCATATGAT 3139
Oy 983 sIleArgAsnPheValValAlaPheLysAsnAsnSerThrTyrLysGln 999
Db 3140 GATATCAGTATTTTCAAAATTTTAAAGAAATTCGAATCAAAATAATCAG 3188

RESULT 14
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DOUBRESTES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Alignment Scores:
Pred. No.: 4,17e-10 Length: 5361
Score: 186.50 Matches: 219
Percent Similarity: 35.09% Conservative: 181
Best Local Similarity: 19.21% Mismatches: 390
Query Match: 3.51% Indels: 351
Gaps: 49
DB: 4

US-09-515-363c-2 (1-1025) x US-08-973-462-2 (1-5361)
Oy 42 ValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValIleLeu 61
Db 967 GTTGAAGAAAGTATGCTGAAATGTTCAAGAAAGTATGCTGAAATGTTGAATAATG 1026
Oy 62 LeuLeuSerThrLeuGluLysGlyValIleThrPheLysLeuGlyTrpThrArgGluIleValGln 81
Db 1027 GTAGCTCAACGCTTCAAAATCGTA----- 1053
Oy 82 AlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluIleIleAsp 101
Db 1054 -----GCTCCAACTGTTGAAGAAATGTTGACGTCGAAGTGTGTTGAGA 1095
Oy 102 LeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGln----- 116
Db 1096 ACTGTGCTCCAAAGTGTGAGAAAGTGTGAGAAAGTGTGAGAAAGTGTGAGAA 1155
Oy 116 ----- 116
Db 1156 AATGTTGAAGAAAGTGTAGCTGAATAATGTTGAGAAAGTGTGCTGAAATGTTGAGAA 1215
Oy 117 -----LeuLeuAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 123
Db 1216 AGTGTAGCTGAAGAAATGTTGAAGAAAGTGTAGCTGAAGAAATGTTGAAGAAATGTTGA 1275
Oy 124 ThrLeu-----ValAspLysLeuLeuValArgAspValIleLeuAsp 136
Db 1276 ACTGTTGAAGAAATGCTAGCTCCAACTGTTGAAGAAATGTTGAGTCCAAAGTGT 1335

```

QY 137 LysCysMetGluGluGluLeuThrIleGluAspArgAsnArgIleAlaIleGlu 156
 Db 1336 AGTGGCTCCAACTGTTGAAGAAAGTGTGAAGAAATCTTGAGAAAGTGTAGTGA 1495
 QY 157 AsnAsnGlyAsnGlyValArgIleLeuLeuLysArgIleValGluLysGluAsn 176
 Db 1396 AAT-----GTTGAAGAAAGTGTAGTGAAGAAATCTTGAGAAAGTGTAGTGA 1449
 QY 177 ThrPheSerAlaPheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGluGlu 196
 Db 1450 GAACAAAGTGTAGTGAAGAAATCTTGAGAAAGTGTAGTGAAGAAATCTTGAGAA 1503
 QY 197 LeuThrGlySerAspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAsp 216
 Db 1504 AGTGTACTGAAATGTTGAAGAAAGTGTAGT-----GAAATGTTGAAGAAATCTGA 1557
 QY 217 GlyProGlnValGluGluGluLeuLeuSerThr----- 227
 Db 1558 GCTCCAACTGTGAAAGAAATCTGAGTCCAACTGTGAAAGAAATCTGAGTCCAACTGT 1617
 QY 228 -----ThrValGlnProAsnLeuGluLysGluValIleProLysGluAsnSerSer 245
 Db 1618 GTAGAAAGTGTGGCTCCAACTGTGAAAGAAAGTGT-----GAAAGAAATCTTGAA 1668
 QY 246 GluSerSerPheAlaAspSerSerValSerGluSerAspThrSerLeuAlaGluLys 265
 Db 1669 GAA-----AGTGTAGTGAAGAAATCTTGAGAAAGTGTAGTGAAGAAAT 1710
 QY 266 SerValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMetGlySerAspSerGly 285
 Db 1711 -----GTTGAAGAAAGTGTAGTGAAGAAATCTT----- 1737
 QY 286 ThrMetClySerAspSerAspGluAsnValAlaAlaArgAlaSerProGluProGlu 305
 Db 1738 -----GAAAGAAAGTGTAGTGAAGAAATCTTGAGAAAGTGTAGTGAAGAAATCT 1788
 QY 306 LeuGlnLeuArgProGlyGlnMetGluValAlaGlnProAlaLeuGluLysAsnIle 325
 Db 1789 GAAATCTGACTCCAACTGTGAAAGAAATCTGCTCCAACTGT-----GTA 1836
 QY 326 IleIleCysLeuProThr-----GlySerGlyLysThrArgAlaValIleAla 343
 Db 1837 GAAAGTGTGGCTCCAACTGTGAAAGAAAGTGTGAAGAAATCTTGAGAAAGTGTAGT 1896
 QY 344 LysAspIleIleuAspLysLysLysAlaSerGluProGlyLysValIleValLeuVal 363
 Db 1897 GAA-----AATGTTGAAGAAAGTGTAGTGAAGAAATCTTGAGAAAGTGTAGTGA 1947
 QY 364 AsnLysValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProIleLeuLys 383
 Db 1948 AATGTTGAAGAAATCTGACTCCAACTGTGAAAGAAATCTGACTCCAACTGTGAGAA 2007
 QY 383 s-----TrpTrpArgValIleGlyLeuSerGlyAspThr 394
 Db 2008 ATTGACTCCAACTGTGTAGAAAGTGTGGCTCCAACTGTGAGAAATCTTGAGAAATCT 2066
 QY 394 GlnLeuLysIleSerPheProGluValIleLysSerCysAspIleIleIleSerThrAl 414
 Db 2067 AATGTTGAAGAAAGTGTAGTGAAGAAATCTTGAGAAATCTTGAGAAATCTTGAGAA 2111
 QY 414 AsnIleLeuGluAsnSerLeuLeuAsnLeuGluAsnGlyLysAlaGluIleValGlnLe 434
 Db 2112 TCMAAAATGTTGAAGAAAGTGA-----GCTGAAATGTTGAAGAAAGTGTAGTGA 2165
 QY 434 userAsp----- 436
 Db 2166 TGTGGAAGAAATCTGACTCCAACTGTGAAAGAAATCTGACTCCAACTGTGAGAAAT 2225
 QY 437 -PheSerLeuIleIleIleAspGluCysHisIleThrAsnLysGluAlaValIleTyrAsn 456
 Db 2226 TGTAGCTCCAACTGTGTGAAAGAAAGTGTGGCTCCAACTGTGTGAAAGAAAGTGTGA 2285

QY 456 nile-----MetArgHisIleuMetGlnLysLeuLysAsnAs 469
 Db 2286 TGTGGAAGAAAGTGTAGTCCAAATCTTGAGAAAGTGTAGTGAAGAAATCTTGAGAA 2345
 QY 469 nArgLeuLysGluAsnLysProValIlePro-----LeuProGlnIleLeuGlnProVal 488
 Db 2346 TGTAGCTGAAGAAATGTTGAAGAAAGTGTAGTCCAACTGTGAGAAAGTGTAGTGA 2405
 QY 488 AlaSerProGlyValGlyValAlaThrLysGlnAlaLysAlaGluGluHisIleLeuLys 508
 Db 2406 TGTGGAAGAAAGTGTAGTCCAACTGTGAGAAAGTGTAGTGAAGAAATCTTGAGAA 2465
 QY 508 sLeuCysAlaAsnLeu----- 514
 Db 2466 TTTATGAGCAATCTTTATAGTAATTTATAGTGTATAGTGAAGTGAAGAAATGAAGA 2525
 QY 514 AlaPhe-----ThrIleLysThrValLysGluAsn-----LeuAspGlu 527
 Db 2526 CAGTATATTAAATGAGATAGAAAGTGAAGAAATGTTAGTCCAACTGTAGTGAAGAA 2585
 QY 527 nLeuLysAsnGlnIleGlnGluProCysLysLysPheAla-----IleAlaAspAlaThrAr 546
 Db 2586 CGTAGAAGAAAGTACAGTGAAGTGTAGTACTACTTTAGTACATATTAGACAGAAATGA 2645
 QY 546 gGlu-----AspProPheLysGluLysLeuLeuGluIle----- 557
 Db 2646 AGAAATACTATTACTAATAGTACTATAGAGAAATTTAGAGAAAGTCCAACTGAATGT 2705
 QY 558 -----MetThrArgIleGlnThrLysCysGlnMetSerPrometSerAspPh 573
 Db 2706 ATTAAGTGGCGCTTGAAGAAATACCCAAAGTGAAGAGAAAGAAAGAAAGTGAAGTGT 2765
 QY 573 eGlyThrGlnProGlyGluGlnIlePalaIle-----GlnMetGlu 586
 Db 2766 AATTGAAGAAAGTGAAGAAAGTGTAGTCCAACTTTAATAGAAAGTGTGAAGAAATGA 2825
 QY 586 uLysLysAlaAlaLysLys-----GlyAsnAlaGlyLysGluArgVal 599
 Db 2826 AGAAAGAGCGCAATATACCAATAGCGAAATATTGAAATATTGAAAGAAAGTGAAGA 2885
 QY 600 -----CysAlaGluHisLeuArgLysTyrAsnGluAlaLeu----- 611
 Db 2886 AAGTATGAAATTTGTCGAGCAATTTAGCAAAATTTAAACGAACTGATTATATATGT 2945
 QY 612 -----GlnIleAsnAspThrIleArgMetIleAspAlaIleThrHisLeuGlnIlePhe 629
 Db 2946 ATTGATTAAGTAGAGCAAGCAAGTGAAGAAATTTAGCGCA-----GAAAGTGT 2990
 QY 629 eTyrAsnGluGluLysAspLysLysPheAlaValIleGluAspAspSerAspGluGlu 649
 Db 2991 AGAAAGCAATGAATGATGAAGCAATTTTATAGTGAATATTGTATTAATGTAAGAAAT 3050
 QY 649 yAspAspGluTyrCysAspGly----- 657
 Db 3051 ACAAGAAATTTATTAACAGATATGTTGGAAGTATAGAAACCATATAGTAATATATC 3110
 QY 657 pGluAspGluAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheGlnMet 677
 Db 3111 ACAAGAAAGGTTGATTTG----- 3129
 QY 677 tThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluIleArg 697
 Db 3130 -----AATGAAATGTTGTTAGTTCGATTTTATGATTAATATAGAAATAT 3173
 QY 697 uAsnGlu-----LysLeuThrLysLeuArgAsnThrIleMetGluGlnIlePheGlnIle 716
 Db 3174 GAAAGAGCTTATTATTAATTAATTAAGAAATTT----- 3207
 QY 716 uGluSerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaIleArgIleGln 736
 Db 3208 -TCAAGTACTGAAGTGTTCAGAAAGTGTAGTCAAGT----- 3246
 QY 736 nTrpIleThrGluAsnGluLysPheAlaGluValGlyValLysAla-----HisIle 754

Db 1955 -----GTTGAAGAAAGTAGCTGAAATCTT----- 1981
 Oy 286 ThrmelYserAspSerAspGluGluAsnValAlaAlaArgAlaSerProGluProGlu 305
 Db 1982 -----GAGAAAGTAGCTGAAATGTTGAAAGAAATCGTAGCTGACTGTTGAA 2032
 Oy 306 LeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGluValAsnIle 325
 Db 2033 GAATGCTAGCTCACTGTTGAAAGAAATGTTAGCTCCAGTCT-----GTA 2080
 Oy 326 IleIleCysLeuProThr-----GlySerGlyThrArgValAlaValTyrIleAla 343
 Db 2081 GAAAGTGGCTCCCAAGTGTGAAAGAAAGTAGAACAAATGTTGAAAGAGTAGCT 2140
 Oy 344 LysAspHisLeuAspLysLysLysLysAlaSerGluProGluValIleValLeuVal 363
 Db 2141 GAA-----AATGTTGAGAAAGTAGTAGCAAAATGTTGAGAAATGTTGAGAA 2191
 Oy 364 AsnLysValLeuLeuValGlu-GlnLeuPheArgLysGluPheGlnProIleLeuLysIle 383
 Db 2192 AATGTTGAGAAATCGTAGCTGCTCACTGTTGAAAGAAATGTTAGCTCCAAATGTTGAGAA 2251
 Oy 383 s-----TPTyrArgValIleGlyLeuSerGlyAspThr 394
 Db 2252 ATTGAGCTCCCAAGTGTGAGAAAGTAGCTCCCAAGTGTGAGAAATGTTGAGAA 2310
 Oy 394 GlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleIleSerThrAl 414
 Db 2311 AATGTTGAGAAAGTAGCTGCTCACTGTTGAAAGAAATGTTGAGAAATGTTGAGAA 2155
 Oy 414 agnIleLeuGluAsnSerLeuLeuAsnGluAsnGlyLysAlaValIleGluIle 434
 Db 2356 TGAAGTGTGAGAAAGTAGCTGCTCACTGTTGAAAGAAATGTTGAGAAATGTTGAGAA 2409
 Oy 434 userAsp-----GCTGAAATGTTGAGAAAGTAGCTGCTGAGAA 2436
 Db 2410 TGTGAGAAAGTAGCTGCTCACTGTTGAGAAAGTAGCTGCTCACTGTTGAGAAAT 2469
 Oy 437 -PheSerLeuIleIleIleAspGluSerHisIleThrAsnLysGluAlaValIleTyrAsnAs 456
 Db 2470 TGTAGCTCCCAAGTGTGAGAAAGTAGCTGCTCACTGTTGAGAAAGTAGCTGCTGAGAA 2529
 Oy 456 nile-----MetArgHisTyrLeuMetGlnLysIleGluLysAsnAs 469
 Db 2530 TGTGAGAAAGTAGCTGCTCACTGTTGAGAAAGTAGCTGCTCACTGTTGAGAAAT 2589
 Oy 469 natGluLysLysGluAsnLysProValIlePro-----LeuProGlnIleGluGlyLeuThr 488
 Db 2590 TGTAGCTGAAAGTAGCTGAGAAAGTAGCTGCTCACTGTTGAGAAAGTAGCTGCTGAGAA 2649
 Oy 488 fAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLysAlaGluGluHisIleLeuLys 508
 Db 2650 TGTGAGAAAGTAGCTGCTCACTGTTGAGAAAGTAGCTGCTGAGAAATGTTGAGAA 2709
 Oy 508 sLeuCysAlaAsnLeu-----As 514
 Db 2710 TTATGAGCAAAATCTTTAGTAATATTATTAGTGTATGCAAACTGAGCAAAATAAGCA 2769
 Oy 514 palAspHe-----ThrIleLysThrValLysGluAsn-----LeuAspGlu 527
 Db 2770 CAGTATATTAAATGAGATGAGAAAGTAGTAAGAAATGTTAGTACCACCAATACTAGCAAAA 2829
 Oy 527 nleuLysAsnGlnIleGlnGluProCysLysLysPheAla-----IleAlaAspAlaThrArg 546
 Db 2830 CCGTAGAAGAAAGTAGCTGAGAAAGTAGTAATCTTTAGTAATATTAGAGAGAGATACA 2889
 Oy 546 gglu-----AspProPheLysGluLysLysLeuLeuGluIle----- 557
 Db 2890 AGAAATATTACTATTACTAATGATAGTAGAGAAATATTAGAGAAAGTAGCTGAGCAAAATGT 2949
 Oy 558 -----MetThrArgIleGlnIleThrTyrCysGlnMetSerProMetSerAspPh 573
 Db 2950 ATTAAGTGGCGCTTGAAGAAATATCCCAAGGTGAGAGGAAAGAAAGAAATTAATGATGCT 3009

Oy 573 eGlyThrGlnProTyrGluGlnTrpAlaIle-----GlnMetGlu 586
 Db 3010 AATTGAGAAAGTAGTAAGAGAGAGTGGCTTACCACTTAATAGAAAGTAGTGAGCAAGAGCA 3069
 Oy 586 uLysLysAlaAlaLysLys-----GlyAsnAlaGlyGluArgVal----- 599
 Db 3070 AGAAAGAGCGCAAAATACAAATTACGCAAAATATTGAAATATTGAAAGAAATGCAATTAGA 3129
 Oy 600 -----CysAlaGluHisLeuArgLysTyrAsnGluAlaLeu----- 611
 Db 3130 AAGTAAATGAATGCTGAGAGATTTAGCAAAATTAAGCAAGCGTATTATTAAATGCT 3189
 Oy 612 -----GlnIleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGlnIleThrPh 629
 Db 3190 ATTAGATAAGAGAGAGCAAGCAAGTAAATTTAGCGCA-----GAAAGCTT 3234
 Oy 629 eTyrAsnGluGluLysAspLysLysPheAlaValIleGluAspAspSerAspIleGlyG 649
 Db 3235 AGAAAGCAATGAAATGATTAAGCATTTTGTAGTAAATATTGATTAATGATTAATGATTA 3294
 Oy 649 yAspAspGluTyrCysAspGly-----As 657
 Db 3295 ACAAGAAATTTATTACAGATATGTTTGCAGATATAGAAACCATATAGTAATTAATG 3354
 Oy 657 pGluAspGluAspAspLeuLysLysProLeuLysLeuAspGluThrAspArgThrPhePhe 677
 Db 3355 ACAAGAAAGTAGTGAATTTG----- 3373
 Oy 677 thrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluIleG 697
 Db 3374 -----AATGAAATGCTGTTAGTGGTCACTTTAGTAAATATAGAAATAT 3417
 Oy 697 uAsnGlu-----LysLeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgIleHrgl 716
 Db 3418 GAAAGAGTATTATTAAATTAATTTAGAAATTT----- 3451
 Oy 716 uGluSerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaIleGlnHrgl 736
 Db 3452 -TCAAGTACTGAGAGTGTGCAAGAAAGTAGTAAGTCAAGAT----- 3490
 Oy 736 nTrpIleThrGluAsnGluLysPheAlaGluValGlyAlaLysAla-----HisHisLeu 754
 Db 3491 -----GTAGAACAAATGATATATGATGATGATGATGCTGCTGATGAAGATTAAT 3543
 Oy 754 uIleGly-----AlaGlnHisSerSerGluPhe----- 763
 Db 3544 TTTAGCAATATTAAATGAGCGAGAGGCTTGAAAGAAATGTTTAAATTTGAAATATG 3603
 Oy 764 -----LysProMetIleHrgl 768
 Db 3604 ATTAAAGTGAAGTAGTGAATTAAGTGTAGAGAAATTAAGATGAACCGGTAAAA 3663
 Oy 768 nasnGluGluGluValIleSerLysPheArgThrGlyLysIleAsnLeuIleAla 788
 Db 3664 AGAGTAGCAAAAGAAAGTAGTATTATTGAGAAATGCAAGCAAAAT-----ATTGT 3717
 Oy 788 aThrThrValAlaGluGluGluLysAspIleLysGlnCysAsnIleValIleAsnGlyG 808
 Db 3718 AGATGATTAGAGCAAGAAAGAGATTAAAGAC----- 3754
 Oy 808 yLeuValThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluIle 828
 Db 3755 -----AAGATGATATGATGCA-----GTAGCAAAATATC 3780
 Oy 828 rThrThrValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAs 848
 Db 3781 CATAGCAATA-----TCTTCAGATTCTTAAGCAAGAAAGTAGATCTATTAAAGA 3828
 Oy 848 pPheArgGluLysMetMetTyrLysAlaIleHisCysValGlnAsnMetLysProGluG 868
 Db 3829 T-----AAAGAAAGAGATGTTTCACTAGTGTGTAAGAAAGTCAAGCAATGATATGATGA 3885

```
QY 868 uTyraIaHlLysIlLeuGlLeuGlMetGlSerIleMetGlLysIsmetLys-- 887
      |||:|||||
Db 3886 AAGTGTGAGAAAGTTTACAAATTG-----AAAAATATGCAAGAGAGAGTAAATGAGCA 3939
      |||:|||||
QY 887 ----- 887
Db 3940 TGCCTGTGAATAATGACATTAAGCAAACTTATGAGAAACTCAGAGTTAAATGA 3999
QY 888 -ThrlYsArGAsnIlleAlaLysHlStYrLysAsnAsnProSerLeuIleIhrPheLeuCy 907
      |||:|||||
Db 4000 AGTAGAAGCAAGATTAAATAAGCATATGCAAAATTAAGAAATTGCAAAAGACATTATC 4059
      |||:|||||
QY 907 sLysAsnCysSerValLeuAlaCysSerGlYlUaspIle-----HisValIleGlulY 925
      |||:|||||
Db 4060 AGAAGATTCTAAAGAAATTAATAGATGCAAAAGATGATACATTAGAAAAGTATATGAGCA 4119
      |||:|||||
QY 925 smethIshIValAsnmetThr---ProGlupheLysGluleuYrIleValArgGIuas 944
      |||:|||||
Db 4120 GGAACATGATATACGACGCGTGGATGAAGTTGTAGAAATTAAGATGCGAAGAGAGA 4179
      |||:|||||
QY 944 nLysAlaLeuGlLysLysCysAlaAspYrGlInIleAsnGlYlUleIleCysLys 963
      |||:|||||
Db 4180 CAAAGATC-----GAAAAAGTATCTGATTTAAAGATCTTGACAGAGATA ATTAAAA 4231
      |||:|||||
```

Search completed: May 16, 2003, 16:35:43
Job time : 328 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 10:02:59 ; Search time 593 seconds

(without alignments)
6539.089 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311
Sequence: 1 MSNGYSTDENRYLISCFRA.....LPITFPNDYSEGLFSDSD 1025

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame_p2n.model -DEV-x1h
-O/-cgn2_1/USPTO-SPool/US09515363/runat_16052003_100245_24779/app_query.fasta.1.1223
-DB-N.Geneseq.101002 -OFMT-fastap -SUFFIX-frag -MINMATCH-0.1 -LOOPEXT-0
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-Biosum62 -TRANS-human40.cdi
-LIST-45 -DOCALIGN-200 -THR-SCORE-pct -THR-MAX-100 -THR-MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFMT-pto -NORR-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
-USER-US09515363.ecgn2_1.1223 -runat_16052003_100245_24779 -NCPU-6 -ICPU-3
-NO_XLPTX -NO_MAP -LARGEOUTERY -NEG-SCORES-0 -WA-T -LONLOG -DEV-TIMEOUT-120
-NARN-TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -Fgapop-6 -Fgapext-7
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

N.Geneseq.101002: *
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001.DAT: *
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: *
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 5311 | 100.0 | 3365 | AA017203 | Human melanoma dif |
| 2 | 5285 | 99.5 | 3372 | ABA04908 | Human RNA helicase |
| 3 | 5276.5 | 99.4 | 3131 | AA011170 | Human melanoma dif |
| 4 | 3176 | 59.8 | 1967 | ABA04960 | CDNA encoding nove |
| 5 | 2442 | 46.0 | 1443 | ABA04916 | Human RNA helicase |
| 6 | 2356.5 | 44.4 | 1557 | AA159285 | Human polynucleoti |
| 7 | 2120 | 39.9 | 1382 | AA161071 | Human polynucleoti |
| 8 | 1995 | 37.6 | 1284 | ABA04913 | Human RNA helicase |
| 9 | 1961.5 | 36.9 | 1319 | AA041517 | CDNA encoding nove |
| 10 | 1344 | 25.3 | 2613 | AAH15201 | Human CDNA sequenc |
| 11 | 1201.5 | 22.6 | 1258 | AA01149 | Interferon induced |
| 12 | 1201.5 | 22.6 | 1270 | AA01151 | Interferon induced |
| 13 | 1106.5 | 20.8 | 956 | AA091688 | CDNA encoding nove |
| 14 | 991 | 18.7 | 1978 | AA040969 | CDNA encoding nove |
| 15 | 968.5 | 18.3 | 3692 | AA011150 | Interferon induced |
| 16 | 969.5 | 18.3 | 3704 | AA01152 | Interferon induced |
| 17 | 830.5 | 15.6 | 1346 | AA041524 | CDNA encoding nove |
| 18 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 19 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 20 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 21 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 22 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 23 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 24 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 25 | 783 | 14.7 | 499 | AA073600 | Human foetal liver |
| 26 | 782 | 14.7 | 609 | AA033745 | Human colon cancer |
| 27 | 782 | 14.7 | 609 | AA033745 | Human colon cancer |
| 28 | 708 | 13.3 | 3026 | AB054189 | Human ovarian anti |
| 29 | 666.5 | 12.5 | 1600 | AA018139 | Human CDNA sequenc |
| 30 | 666 | 12.5 | 1606 | AA018139 | Human CDNA sequenc |
| 31 | 666 | 12.5 | 1606 | AA018139 | Human CDNA sequenc |
| 32 | 649 | 12.2 | 374 | AA017207 | Human melanoma dif |
| 33 | 642 | 11.9 | 392 | AA069625 | EST clone COS22. |
| 34 | 630 | 11.9 | 387 | AA069625 | Novel human polyinu |
| 35 | 563.5 | 10.6 | 3184 | AA064495 | Human ovarian polyu |
| 36 | 553 | 10.4 | 340 | AA144761 | Human breast cance |
| 37 | 553 | 10.4 | 341 | AA123623 | Human breast cance |
| 38 | 498 | 9.4 | 458 | AA046601 | Nucleotide sequenc |
| 39 | 488 | 9.2 | 292 | ABA07057 | Human breast cell |
| 40 | 488 | 9.2 | 292 | ABA04940 | Human foetal liver |
| 41 | 488 | 9.2 | 292 | ABA04940 | Probe #10515 for g |
| 42 | 488 | 9.2 | 292 | AAK13367 | Human brain expres |
| 43 | 488 | 9.2 | 292 | AAK13367 | Human brain expres |
| 44 | 488 | 9.2 | 292 | AAK13367 | Human bone marrow |
| 45 | 488 | 9.2 | 292 | AA15913 | Probe #9846 for qe |
| | | | | AA15108 | Probe #13794 used |

ALIGNMENTS

RESULT 1

ID AAD17203 standard; CDNA: 3365 BP.

AC AAD17203;

DT 29-NOV-2001 (first entry)

XX Human melanoma differentiation associated (Mda)-5 cDNA.

DE Human: melanoma differentiation associated gene; Mda-5; interferon; IFN;

XX RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;

KW neuroblastoma; astrocytoma; glioblastoma; multiforme; cancer; cervical;

KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;

KW central nervous system; cytostatic; apoptosis; ss.

OS Homo sapiens.

XX

Db 1655 AAGGCAAGACACATTTTAAATATGTCATCTGATGACATTTAATTAAGCT 1714
 Oy 521 ValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnIleProCysLysPheAla 540
 Db 1715 GTTAAAGAAACCTTATCATGAGAAAAACCAATACAGAGCATGACAGAGATTGGC 1774
 Oy 541 IleAlaAspAlaThrArgGluAspPropheLysGluLysLeuGlnIleMetThrArg 560
 Db 1775 ATTGCAGATGCCAACAGACAGATCATTTAAAGAACCTTCTACAAATATGACAAAG 1834
 Oy 561 IleGlnThrTyrCysGlnMetSerPrometSerAspPheGlyThrGlnProTyrGln 580
 Db 1835 ATCAAACTTATGTCAAATGAGTCCATGCAATGTCAGATTGGAACTCAAGCTATGAACA 1894
 Oy 581 ThrAlaIleGlnMetLysLysAlaIleLysGlyAsnArgLysGluArgValLys 600
 Db 1895 TGGGCAATTCAAATGAGAAAAAGCTGCAAAAAGAGAAATCCAAAGCAAGCTGTCT 1954
 Oy 601 AlaGlnIleLeuArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleArgMetIle 620
 Db 1955 GCGAAGACATTTGGAGAGTACAAATTAAGCCCTTCAAAATTAATACCAATTCGAATGATA 2014
 Oy 621 AspAlaTyrThrIleLeuGlnThrPheTyrAsnGlnLysAspLysPheAlaVal 640
 Db 2015 GATGCGTATCATCTTGAACTTCTCTAATGAGAGAGAAATAGAGATTTGGCAGTC 2074
 Oy 641 IleGluAspSerAspGluGlyGlyAspAspGluTyrCysAspGlyArgGluAspGlu 660
 Db 2075 ATGGAAGATCATGTGATGAGGGGTGGTGAATGATGATGTGATGATGATGATGATGAT 2134
 Oy 661 AspAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhe 680
 Db 2135 GATGATTTTAAAGAAACCTTTGAACTGATGATGATGATGATGATGATGATGATGAT 2194
 Oy 681 PheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProLysLysAsnGluLys 700
 Db 2195 TTGGAAGAAATTAATGTTGAAAGAGCTGGCTGAAAGAACCAATATGCAATGAAAG 2254
 Oy 701 LeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgThrGlnIleSerAlaArg 720
 Db 2255 CTGACCAAAATTAAGAAATACCAATATGAGCAATATGAGCAATGAGCAATGAGCAAT 2314
 Oy 721 GlyIleIlePheThrLysThrArgGlnSerAlaThrAlaLeuSerGlnThrIleThrGlu 740
 Db 2315 GGAATATCTTTACAAAAACACACAGAGTGCATATGCGTTTCCAGTGCATATCTGAA 2374
 Oy 741 AsnGluLysPheAlaGluValGlyValLysAlaIleHisLeuIleGlyAlaGlyHisSer 760
 Db 2375 AATGAAAAATTTGCTGAGTAGAGATCAAGCCACCATCTGATGAGATGGACACAGC 2434
 Oy 761 SerGluPheLysPrometThrGlnAsnGluGlnLysGluValIleSerLysPheArgThr 780
 Db 2435 AGTGAGTTCAAACCATGACACAGCAATGAAACAAAGAAAGATCATAGTAATTTCCGAT 2494
 Oy 781 GlyLysIleAsnLeuLeuIleAlaThrThrValAlaGlnGluGluLeuValIleLysGlu 800
 Db 2495 GGAATAATTAATCTGCTTATCGCTACACAGCTGGCAGAGAGAGCTTGGAATTAAGAA 2554
 Oy 801 CysAsnIleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArg 820
 Db 2555 TGTAACATGTGTATCGGTATGCTGCTGTCACCAATGAATAGCATAGTCAGGCGGT 2614
 Oy 821 GlyArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlySerGlyVal 840
 Db 2615 GGTGAGCCAGAGCTGTATGAGCACCTACGCTCTGCTGCTCAAGTGTTCAGAGATT 2674
 Oy 841 IleGlnIleGlnThrValAsnAspPheArgGluLysMetMetTyrLysAlaIleHisCys 860
 Db 2675 ATGCAACGTAGACAGTAAATGATTTCCGAGAGACATGATGATAAAGATACATTTGT 2734
 Oy 861 ValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGlnLeuGlnMetGlnSer 880

Db 2735 GTTCAAAATATGAAACACAGAGAGTATGCTCATAGATTGGATTACAGATTAAGT 2794
 Oy 881 IleMetGluLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAspPro 900
 Db 2795 ATAAATGAAAGAAATGAAATGAAACCAAGAGAAATATTTGCCAAGCATTAACAGATTAACCA 2854
 Oy 901 SerLeuIleThrPheLeuLysLysAsnGlySerValIleAlaCysSerGlyLysAspIle 920
 Db 2855 TCACATAATACCTTCTCTTTCGAAAAAGTCGACATGTGTAGCTGTTCGGGCAATATATC 2914
 Oy 921 HisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLysIle 940
 Db 2915 CATGTAATGAGAAAAATGATCAGCTCAATATGACCCAGAAATTCAGAGAACTTTTCAAT 2974
 Oy 941 ValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyThrIle 960
 Db 2975 GTTACGAGAAACCAAGACATGCAAAAGAGTGTGCGATATCAAAATTAATGATTAATC 3034
 Oy 961 IleCysLysCysGlyGlnAlaThrPheGlyThrMetMetValHisLysGlyLeuAspPhePro 980
 Db 3035 ATCTGCAATGTGGCCAGGCTTGGGGAACATGATGTGCAAAAGGCTTACATTTCCCT 3094
 Oy 981 CysLeuLysIleArgAsnPheValValValPheLysAsnAsnSerThrLysLysGlnIyr 1000
 Db 3095 TGCTCAAAATTAAGGAATTTTGTAGTGTTCGAAATTAATTCAAACAAAGAAACAAATAC 3154
 Oy 1001 LysLysThrValGluLeuProIleThrPheProAsnLeuAspTyrSerGluCysLysLeu 1020
 Db 3155 AAAAAGTGGTGAATTAATCAATCAATTCATTCATTCATTCATTCATTCATTCATTCAT 3214
 Oy 1021 PheSerAspGluAsp 1025
 Db 3215 TTTAGTATGAGAGAT 3229
 RESULT 3
 AAD11170
 ID AAD11170 standard; DNA; 3131 BP.
 XX
 AC AAD11170;
 XX
 XX 29-NOV-2001 (first entry)
 DT
 XX
 XX Human melanoma differentiation associated-5 protein-related DNA.
 DE
 KW Human melanoma differentiation associated gene; Mda-5; Interferon; IFN;
 KW RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;
 KW neuroblastoma; astrocytoma; glioblastoma; multifactor; cancer; cervical;
 KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
 KW central nervous system; cytosolic; apoptosis; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200164707-A1.
 PN
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06960.
 XX
 PR 29-FEB-2000; 2000US-051363.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Fisher PB, Kang D, GopalKrishnan RV;
 XX
 DR WPI; 2001-565494/63.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity -
 XX
 PS Disclosure: Page 134-148; 152pp; English.
 CC The present invention relates to an isolated nucleic acid encoding a

PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226688.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227209.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229509.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231245.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 12-SEP-2000: 2000US-0231968.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0235484.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235835.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.
PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239935.
PR 13-OCT-2000: 2000US-0239937.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241221.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.
PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 20-OCT-2000: 2000US-0241826.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0244674.
PR 08-NOV-2000: 2000US-0244675.
PR 08-NOV-2000: 2000US-0244676.
PR 08-NOV-2000: 2000US-0244677.
PR 08-NOV-2000: 2000US-0244678.
PR 08-NOV-2000: 2000US-0245523.
PR 08-NOV-2000: 2000US-0245524.
PR 08-NOV-2000: 2000US-0245525.
PR 08-NOV-2000: 2000US-0245526.
PR 08-NOV-2000: 2000US-0245527.
PR 08-NOV-2000: 2000US-0245528.
PR 08-NOV-2000: 2000US-0245532.

PR 08-NOV-2000: 2000US-0246609.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0251989.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251866.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-465566/50.
P-PSDB: AAU23090.

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases

Claim 4: SEQ ID No 186; 1180bp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU3814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. AA540785-AA541684 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1967 bp; 723 A; 344 C; 416 G; 476 T; 8 other;

Alignment Scores:

| Prod. No.: | 1.2e-252 | Length: | 1967 |
|------------------------|----------|---------------|------|
| Score: | 3176.00 | Matches: | 636 |
| Percent Similarity: | 97.11% | Conservative: | 2 |
| Best Local Similarity: | 96.80% | Mismatches: | 18 |
| Query Match: | 59.80% | Indels: | 5 |
| DB: | 22 | Gaps: | 1 |

US-09-515-363c-2 (1-1025) x AAS40960 (1-1967)

```

QY 294 GluAsnValAlaAlaArgAlaSerProGluLeuGlnLeuArgProGlyGlnMet 313
    |||||
DB 3 GAGAAATGGCCAGCAGACAGATCCCGGAGCCAGAACTCAGCTTACAGGCTTACCAATG 62
    |||||

QY 314 GluValAlaGlnProAlaLeuGlnGlyLysAsnIleIleCysLeuProThrGlySer 333
    |||||
DB 63 GAAGTGGCCAGCAGCAGCTTGGAGGAAATATCATCTGCTCCCTACAGGAGT 122
    |||||

QY 334 GlyLysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLysLysLysAla 353
    |||||
DB 123 GGAATAACCAAGAGTGGCTGTTTACATTCAGCAGATCCTTAACAAGAGCAAAAAAGCA 182
    |||||

QY 354 SerGluProGlyLysValIleValLeuValAsnLysValLeuValGlnLeuPhe 373
    |||||
DB 183 TGTGAGCTGGAGAAAGTTATAGTCTTGTGCAATAGGTAAGTCTTGTGATAGCTCTTC 242
    |||||

QY 374 ArgLysGluPheGlnProPheLeuLysLysTyrTyrArgValIleGlyLysSerGlyAsp 393
    |||||
DB 243 CGCAAGAGCTCCACCATTTTGGAGAAATGATGCTGCTTAATGATTAAGTCTGAT 302
    |||||

QY 394 ThrGlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleIleSerThr 413
    |||||
DB 303 ACCCACTGAAAAATATCATTCATTCAGAAAGTTGTAAGTCTGTTATATTATATCATGACA 362
    |||||

QY 414 AlaGlnIleLeuGlnLysSerLeuLeuAsnLeuGlnLysGlnLysAlaGln 433
    |||||
DB 363 GCTCAATCTCTTGAATACTCTCTTAACTTGAATAATGAGAGATGCTGCTGCTGCTCA 422
    |||||

QY 434 LeuSerAspPheSerLeuIleIleIleAspGluCysHisHisThrAsnLysGlnAlaVal 453
    |||||
DB 423 TTGTGAGCTTTCCCTCATATATCATGATGATGATGATGATGATGATGATGATGATGATG 482
    |||||

QY 454 TyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeuLysAsnAsnIleGluLys 473
    |||||
DB 483 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 542
    |||||

QY 474 GluAsnLysProValIleProLeuProGlnIleLeuGlnLeuThrAlaSerProGlyVal 493
    |||||
DB 543 GAAAAACAAACAGATGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
    |||||

QY 494 GlyGlyAlaThrLysGlnAlaLysAlaGlnGlnHisIleLeuLysLeuGlnLysAlaLeu 513
    |||||
DB 603 GGAAGGGGCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 662
    |||||

QY 514 AspAlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 533
    |||||
DB 663 CATGCAATTTCTATTAATCTGTTAAAGAAACCTTCATGATTAATGAAGAAACCAATATCAG 722
    |||||

QY 534 GluProCysLysLysPheAlaIleAlaAspAlaThrArgLysAspProPheLysGlnLys 553
    |||||
DB 723 GAGCAGTCAAGAAAGTTGCTGATTCAGATGCAACGAGACATTCATTTAAAGAGAAA 782
    |||||

QY 554 LeuLeuGlnIleMetThrArgIleGlnIleThrTyrCysGlnMetSerProMetSerAspPhe 573
    |||||
DB 783 CTCTTCAAAATATATGACAAAGATTCAAACTTATGTCAAATATGTCAAATATGTCAGATT 842
    |||||

QY 574 GlyThrGlnProTyrGlnGlnIleProAlaIleGlnMetGlnLysLysAlaIleLysGly 593
    |||||
DB 843 GGAAGTCAACCTATGACAAATGAGCCATTCAAATGAAAAAAGAGCTTAAGAGAGAGGA 902
    |||||

QY 594 AsnArgLysLysLysValCysAlaGlnHisLeuArgLysTyrAsnGlnAlaLeuGlnIle 613
    |||||
DB 903 AATCGCAAGAGAGCTGTTGTCAGAAACATTTGAGAGATGATGAGCTTACAAATTT 962
    |||||

```

```

QY 614 AsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGlnThrPheTyrAsnGlnGlu 633
    |||||
DB 963 AATGACACAAATTCGATGATATAGATGCTGATATCTGATCTGTAATATTAAGAG 1022
    |||||

QY 634 LysAspLysLysPheAlaValIleGlnLysAspSerAspGluGlyGlyAspAspIleTyr 653
    |||||
DB 1023 AAAGATTAACAACTTTCGACTCATACAGATGATATGATGAGAGTGGTGGATGATATAT 1082
    |||||

QY 654 CysAspGlyAspGluAspGluAspAspLeuLysLysProLeuLysLeuAspGlnThrAsp 673
    |||||
DB 1083 TGTGATGCTGATGAGATGAGATGATGATTTAAAGAAACCTTGAACCTGATGATGAAACAGAT 1142
    |||||

QY 674 ArgPheLeuMetThrLeuPhePheGlnLysAsnLysMetLeuLysArgLeuAlaLysLys 693
    |||||
DB 1143 AGATTTCTCATGACTTATTTTATTTTAAACAAATTAATTTGTAAGAGCTGCTGCTGCTG 1202
    |||||

QY 694 ProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGlnGlnIleThr 713
    |||||
DB 1203 CCAGATATGAAAAATGAAAAAGCTGACCAATTTAAGAAATACCATATGAGTAATATACT 1262
    |||||

QY 714 ArgThrGlnGluSerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaIleVal 733
    |||||
DB 1263 AGCACTGAGCAATTCAGACAGCAAGAAATATCTTTACAAAAACAGACAGAGTGCATATGCG 1322
    |||||

QY 734 LeuSerGlnThrIleThrGlnLysGlnLysPheAlaGlnValGlyValLysAlaHisHis 753
    |||||
DB 1323 CTTTCCCACTGATTAAGTGAATAATGAAAAATTTGCTGAAGTGAAGTGAAGTGAAGTGA 1382
    |||||

QY 754 LeuIleGlyAlaGlyHisSerSerGluPheLysProMetThrGlnAsnGlnGlnLysGln 773
    |||||
DB 1383 CTGATTGAGAGCTGAGACAGCAAGTGAATTCCTTTACAAAAACAGACAGAGTGCATATGCG 1442
    |||||

QY 774 ValIleSerLysPheArgThrGlyLysIleAsnLeuIleAlaThrThrValAlaGln 793
    |||||
DB 1443 GTCATTAGTAATTTTCGCACTGAAAAATTAATTCGCTATGCTCC-ACACTGCAAGAA 1501
    |||||

QY 794 GluGlyLeuAspIleLysGlnCysAsnIleValIleArgTyrGlyLeuValThrAsnGln 813
    |||||
DB 1502 GAAGGCTGATATTTAAACAAATGATTAATCATGTTATCCGTTATGCTGCTCCCAATGCA 1561
    |||||

QY 814 IleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluSerThrTyrValGlnVal 833
    |||||
DB 1562 ATATCC-ATGCTGCAAGC-CGTGGTCAAGCTAGAGGATGAGAGCACTACTCT-CTGGTT 1618
    |||||

QY 834 AlaHisSerGlySerGlyValIleGlnHisGlnThrValAsnAspPheArgGlnGlnMet 853
    |||||
DB 1619 GCTCACAGCTGTTCAAGAGATTAATGAACTGAGACAGTAAATGATTTCCGACAGAAAGATG 1678
    |||||

QY 854 MetTyrLysAlaIleHisCysValGlnAsnMetLysProGluGlnTyrAlaHisGlnIle 873
    |||||
DB 1679 ATGTATTAAGCTATTC---ATTGCTTCAAAATATGAACAGAGAGATATGCTATATAGAAAT 1735
    |||||

QY 874 LeuGlnLeuGlnMetGlnSerIleMetGlnLysLysMetLysThrLysArgAsnIleAla 893
    |||||
DB 1736 TTGGATTTACAGATGCAAGTAAATGATGAAGAAAGAAATGAAACCAAGAGAAATATTCGC 1795
    |||||

QY 894 LysHisTyrLysAsnAsnProSerLeuIleThrPheLeuCysLysAsnGlySerGlnGln 913
    |||||
DB 1796 AAGCATTAACAGAAATTAACCATCACAATTAATTCCTTCTTGGAAAAACGCGAGTATGCTA 1855
    |||||

QY 914 AlaCysSerGlyGlnAspIleHisValIleGlnLysMetHisHisValAsnMetThrPro 933
    |||||
DB 1856 GCGTGTCTGGGGAGATATCCATGAAATTTGAAAAATTCATTAACCATATATGACCCCA 1915
    |||||

QY 934 GluPheLysGlnLeuTyrIleValArgGluAsnLysAlaLeuGlnLysLys 950
    |||||
DB 1916 GAATTAAGAGACTTACATTTGGAAGAGAAAGAAACCAAGCTGCCAAAGAGAA 1966
    |||||

```

RESULT 5
ABR04916
ID ABA04916 standard; DNA; 1443 BP.
XX

AC ABA04916:
 XX
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH16 related DNA sequence *.
 XX
 KW Human; RH16; RNA helicase; cytosolic; virucide; anti-HIV;
 KW Immunosuppressive; immunostimulatory; antineoplastic; antidiabetic;
 KW antileukemic; osteoporosis; osteoarthritis; hepatotropic;
 KW anti-inflammatory; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 11-MAY-2001: 2001WO-FR01441.
 XX
 PR 11-MAY-2000: 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 XX
 PI Bahr G, Cocude C, Capron A;
 PI WPI: 2002-082898/11.
 DR
 XX New polypeptide, useful for treating and diagnosing cancer or
 PT inflammation, and drug screening, comprises a human polynucleotide
 PT homologous to RNA helicase.
 XX
 XX Example 2: Page 94-95; 114pp; French.
 XX
 CC The present invention relates to human RH16 (see AAM/798). RH16 is a
 CC 116kDa protein and has homology to RNA helicases (DEXH box). RH16 and
 CC its coding sequence are useful for treating cancer acute or chronic
 CC infections (especially by HIV or hepatitis B or C); inherited genetic
 CC diseases (auto)immune diseases (particularly rheumatism, arthritis,
 CC arteriosclerosis, osteoporosis and diabetes, but many others listed)
 CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.
 XX
 SO Sequence 1443 BP; 508 A; 261 C; 313 G; 360 T; 1 other:
 SO
 Alignment Scores:
 Pred. No.: 3 43e-192 Length: 1443
 Score: 2442.00 Matches: 475
 Percent Similarity: 99.388 Conservative: 2
 Best Local Similarity: 98.968 Mismatches: 3
 Query Match: 45.988 Indels: 0
 DB: 24 Gaps: 0
 US-09-515-363c-2 (1-1025) x ABA04916 (1-1443)
 QY 174 LysGluAsnTrpPheSerAlaPheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeu 193
 DB 2 AAGAAAGAGCTGCTCTCATTTCTGAATGTTCTCTCTCAAGAACAGAAATGAACTT 61
 QY 194 ValGlnGluLeuThrGlySerAspCysSerGluSerAsnAlaGlnIleGluAsnLeuSer 213
 DB 62 GTCCAAAGAGCTTACAGGCTCTCATTTCTCAAGAACAGAAATGAACTTATCA 121
 QY 214 GlnValAspGlyProGlnValGluGluGlnLeuLeuSerThrPheValGlnProAsnGlu 233
 DB 122 CAAGTTCATGCTCTCAAGAGCAAGCAACTTCTTCAACACAGTTCATCAAAATCTG 181
 QY 234 GluLysGluValTrpGlyMetGluAsnAsnSerSerGluSerSerPheValAspSerSer 253
 DB 182 GAGAACGAGGCTGTGGGCAATGAGAAATTAATCATCATCAATCTTCTT 241

QY 254 ValValSerGluSerAspThrSerLeuAlaGluGlySerValSerCysLeuAspThrSer 273
 DB 242 GTAGTTTCAGAAATCAGACCAAGTTTGCACAAAGAGAGCTGCTGCTTCAATTAAGT 301
 QY 274 LeuGlnHisAsnSerAsnMetGlySerAspSerGlyThrMetGlySerAspSerAspGlu 293
 DB 302 CTGGACATTAACAGCAACATGGCAGTATTCAGCCACCATGGGAGATGATTACATAGAA 361
 QY 294 GluAsnValAlaAlaArgAlaSerProGluProGluLeuGlnLeuArgProGlyThrMet 313
 DB 362 GAGATGTGGCAGCAAGAGCATCCCGGAGCAGAACTCCAGCTCAGCTTACCAATG 421
 QY 314 GluValAlaGlnProAlaLeuGlnGluGlyAsnIleIleGlyLeuProThrGlySer 333
 DB 422 GAAGTGGCCAGCCAGCCCTTGGAAAGGAAATATATCATCTCTCCCTACAGGGAGT 481
 QY 334 GlyLysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLysLysVal 353
 DB 482 GGAAGAACACAGAGTGGCTTTTACATTGCCAAGATCTCTTAGACACAGCAAAAAAGCA 541
 QY 354 SerGluProGlyLysValIleValLeuValAsnLysValLeuLeuValGluIlePhe 373
 DB 542 TCTGAGCCTGGAAAGTATAGTTCTGTCAATAGTACTAGTGAACAGCTCTTC 601
 QY 374 ArgLysGluPheGlnProPheLeuLysLysTyrTrpArgValIleGlyLeuSerGlyAsp 393
 DB 602 CGCAAGAGAGTCCAAACCATTTTGAAGAAATGATCTCTTATGCTTAACTGATGAT 661
 QY 394 ThrGlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleLeuThr 413
 DB 662 ACCCAACTGAAATATTCATTTCCAGAACTGTGCAAGTCTGTGATATTATTCATGACA 721
 QY 414 AlaGlnIleLeuGluAsnSerLeuLeuAsnLeuGluAsnGlyGluAspAlaGlyValGln 433
 DB 722 GCTCAATCTCTGAAAGTCCCTCTTAACCTGAAATGAGAAAGAGTGGTCTGTTTCA 781
 QY 434 LeuSerAspPheSerLeuIleIleIleAspGluCysHisIleThrAsnLysGluVal 453
 DB 782 TTGTAGACTTTTCTTATATCATCTTATGATATTCATCACCACCAAGAAATGATG 841
 QY 454 TyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeuLysAsnAspArgLys 473
 DB 842 TATATATCATCATGATGATGATTTGATGCAAGATGTAACAAATAGACTCAAGAA 901
 QY 474 GluAsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGlyVal 493
 DB 902 GAAACAAACACGATGATTCCTTCATGATGAGTCACTGAGTCAAGCTTCACTCTT 961
 QY 494 GlyGlyAlaThrGlyGlnAlaLysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeu 513
 DB 962 GGAGGGCCGACGAGCAAGCCAAAGCTGAGCAAGCAATTTTAAACTTGTGCTAACTCT 1021
 QY 514 AspAlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 533
 DB 1022 GATGCATTACTATTAAAGTGTAAAGAAACCTTGATCAACTGAAACCAAAACACAG 1081
 QY 534 GluProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGluLys 553
 DB 1082 GAGCCATGCAAGAAAGTGTGCACTTGCATGCAACCAAGAGATCCATTTTAAACAGAA 1141
 QY 554 LeuLeuGluIleMetThrArgIleGlnThrTyrCysGlnMetSerProMetSerAspPhe 573
 DB 1142 CTCTAGAAATATATACAAAGATTAACCTTATTTCTCAATGAGTCCATGTCATGATTT 1201
 QY 574 GlyThrGlnProTyrGluGlnTrpAlaIleGlnMetGluLysLysAlaAlaLysLysGly 593
 DB 1202 GGAATCTCAACCTATGAAACATAGGGCATTCAAATGAGAAAAAGCTGCAGAAAGAA 1261
 QY 594 AsnArgLysGluArgValCysAlaGlnHisLeuArgLysTyrAsnGluAlaLeuGlnIle 613
 DB 1262 AATCCAAAGAAAGTGTGTCAGAACATTTGAGAGAGTCAATTAAGCCCTTAATAAT 1321
 QY 614 AsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheTyrAsnIleGlu 633

DB 919 TATACATTGTGTCACAAATATGAAACAGAGAGTATGCTCAAGATTTCGAATTACA 978
 OY 877 nmetGlnSerIleMetGluLysLysMetLysThrLysAlaGlnIleAlaLysHisTyrLys 897
 DB 979 GATGCAAAAGATATATGCAAAAGAAATGAAACCAAGACAAATATGCGCAACCATTTACAA 1038
 OY 897 sAaAaAProSerLeuIleThrPheLeuLysLysAsnGlnValLeuIleGlySerGln 917
 DB 1039 GAATTAACCATCTACTAATTAATCTCTTTCGCAAAACATGCAATCTGCTATCTGTTTCG 1098
 OY 917 ygluAaPIleHisValIleGluLysMetHisHisValAsnMetThrProLysPheLysGln 937
 DB 1099 GGAAGATATCCATGATGATATGAGAAATGATCAGTCATATATACCCCAATTCAGCA 1158
 OY 937 uLeuTyrIleValArgGluAsnLysAlaLeuGlnLysGlySerAlaAspIleGlnIleAs 957
 DB 1159 ACTTACATTTGTAGAGAAACCAAACTGCAAAAGAGTGGCGGATATCAATTA 1218
 OY 957 nglYgluIleIleGlyLysGlyGlnAlaIleTyrGlyThrMetValHisLysGlyLe 977
 DB 1219 TGGTGAATCATCTGCAAAATGTCGCGGCGGATGCGGAAATATGTCGACAAAGCTT 1278
 OY 977 uAaPIleProCysLeuLysIleArgAsnPheValValPheLysAsnAsnSerThrLys 997
 DB 1279 AGATTGCTTGTCTCAAAATAGCAATTTTGTAGTGGTTTAAAAATTAATTCACACA 1338
 OY 997 slYsGlnTyrLysLysTrpValGluLeuProIleThrPheProAsnLeuAspTyrSerGln 1017
 DB 1339 GAACAAATACAAAGAGTGGTACAAATTTACCTATACATTTCTCAATCTTACTATTCACA 1398
 OY 1017 ucYsCysLeuPheSerAspGluAsp 1025
 DB 1399 ATGCTGTTTATTAGTAGATGAGAT 1423
 RESULT 7
 AA161071
 ID AA161071 standard; cDNA; 1382 BP.
 AC AA161071:
 DT 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 5060.
 XX Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 XX peripheral nervous system; neuropathy; central nervous system; CNS;
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 XX leukaemia; ss.
 XX Homo sapiens.
 XX MO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000MO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX PA
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Qa, Zhou P, Goodrich R, Dimanac RT;

XX WP1: 2001-442253/47.
 DR P-PSDB: AAM41915.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Claim 1; SEQ ID NO 5060; 10078bp; English.
 CC The invention relates to human nucleic acids (AA157798-AA161469) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 1382 BP: 513 A; 217 C; 285 G; 367 T; 0 other:
 SQ
 Alignment Scores:
 Pred. No.: 1,29e-165 Length: 1382
 Score: 2120.00 Matches: 407
 Percent Similarity: 98.32% Conservative: 2
 Best Local Similarity: 97.84% Mismatches: 5
 Query Match: 39,92% Indels: 2
 DB: Gaps: 0
 US-09-515-363C-2 (1-1025) x AA161071 (1-1382)
 OY 612 GlnIleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheIleAsn 631
 DB 13 CAATTAATGACACAAATTCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 72
 OY 632 GluGluLysAspLysLysPheAlaValIleGluAspAspSerAspGluGlyLysAsp 651
 DB 73 GAAGGAAAGATGAAAGATTTGCACTCATGATGATGATGATGATGATGATGATGATGAT 132
 OY 652 GluTyrCysAspGlyAspGluAspGluAspLysLysProLeuLysLysAspGlu 671
 DB 133 CAGTATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
 OY 672 ThrAsp-ArgPheLeuMetThrLeuPhePheGluAsnAsnLysMet-LeuLysAlaIleuA 691
 DB 193 ACAGATACGATTTCTCATGACTTATTTTGAACAAATTAATTAATTAATTAATTAATTAAT 252
 OY 691 IagluAsnProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGlu 711
 DB 253 CTGAACACCCCTTAATATGAAATGAAAGCTGACCAATTAAGAAATTAATTAATTAATTAAT 312
 OY 711 IATyrThrArgThrGluGluSerAlaArgGlyIleIlePheThrLysThrArgGluSerA 731
 DB 313 AATATACTAGAGCTAGAGAAATCAGACAGCAATTAATTTTCAAAAACCAATTAATTAAT 372
 OY 731 IATyrAlaLeuSerGlnIleThrGluAsnGluLysPheAlaGluValGlyValLysA 751
 DB 373 CATATGCGCTTCCAGAGATTAATGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 432
 OY 751 IahisHisLeuIleGlyAlaGlyHisSerSerGluPheLysProMetThrGluAsnGlu 771
 DB 433 CCCACCATCTGATGAGACTGACACACACAGCTGATTAACCAATTAATTAATTAATTAATTAAT 492
 OY 771 IATyrGluValIleSerLysPheArgThrGlyLysIleAsnLeuLeuIleAlaThrThr 791
 DB 493 AAAAAGAGTCAATTAGTAATTTTGCACCTGCAAAATTAATTAATTAATTAATTAATTAATTAAT 552

QY 791 aLaagluGluglyLeuaspIleLysglucysasnIleValIleArgTyrAlaLeuValP 811
 Db 553 TGCGAGAGAGAGCTGGATATTAAAGATGTAACATGTTATCCGTATACCTGCTGCA 612
 QY 811 hransgluIlealmetvalGlnalaArgglyAraGalaAspGlnSerThrTyrV 811
 Db 613 CCAATGAAATAGCCATGCTCAGGCCCGTGGTGCAGCCAGAGCTGATGAAAGACCTACG 672
 QY 831 alleuValAlaHisSerGlySerGlyValIleGlnHisGluThrValAspPheArgG 851
 Db 673 TCCTGCTGCTCAGAGGTTTCAGAGATTACACATGAGACAGATTAAJATTCCGAG 732
 QY 851 luLysmetMetTyrLysAlaIleHisCysValGlnasmetLysProGlnIleTyrAlaH 871
 Db 733 AGAAGATGATGTATMAAGCATATGCTGTTCTMAANTATGAAACAGAGAGATGCTC 732
 QY 871 lsLysIleLeuGluLeuGlnMetGlnSerIleMetGlnLysLysMetLysThrLysArgA 891
 Db 793 ATAGATTTTGGAAATTACAGATGCAAAAGTATATGAAAGAAATGAATGAAACAGAGA 852
 QY 891 snIleAlaLysHisTyrLysAsnAsnProSerLeuIleThrPheLeuGlyAsnCys 911
 Db 853 ATATTGCCAAGCATACAAAGATTAACCCATCACTAATTAACCTTCCTTGTAACAGCTGA 912
 QY 911 erValleuAlaCysSerGlyLysAspIleHisValIleGlnLysMetHisHisValAsnM 931
 Db 913 GTGTGCTAGCCTGTTCTGGGGAGATATCATGTAATTCMAAATGCAATACCTCAATA 972
 QY 931 etThrProGluPheLysGluLeuTyrIleValArgGluAsnLysAlaLeuGlnLysLysC 951
 Db 973 TGACCCAGAAATTCAGAACATTTACATTTAGAGAAACCAAAACCTGCAAAAGACCT 1032
 QY 951 ysAlaAspTyrGlnIleAsnGlyLulIleLeuGlyLysGlyGlnAlaArgGlyThrM 971
 Db 1033 GTGCCGACTATCAATTAAGTGAATTCATCTGCAATTCCTGCAAGCTTGCGAGACAA 1092
 QY 971 etHeValHisLysGlyLeuAspLeuProCysIleuLysIleArgAsnPheValValValP 991
 Db 1093 TGATGTGTCACAAAGGCTTGATTTGCTGCTGCTCACAATAAGGAATTTTGTAGTGT 1152
 QY 991 heLysAsnAsnSerThrLysLysGlnTyrLysLysTyrValIleGluLeuPheIleThrPheP 1011
 Db 1153 TCAAAATATATTCACAAACAAACAAATACAAAAGTGGTAGAATTAACCAATTCACATTTG 1212
 QY 1011 roAsnLeuAspTyrSerGlyCysCysLeuPheSerAspGluAsp 1025
 Db 1213 CCAATCTTGACTATTCAGAAATGCTGTTATTAGTGTAGAGAA 1255
 RESULT 8
 ID ABA04913 standard; DNA: 1284 BP.
 AC ABA04913:
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH16 related DNA sequence #2.
 XX
 XX Human; RH16; RNA helicase; cytosolic; virucide; anti HIV;
 KW immunosuppressive; immunostimulatory; antirheumatic; antiarthritic;
 KW antileptosclerotic; osteopathic; antidiabetic; hepatotropic;
 KW antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN NC0200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 11-MAY-2001; 2001MO-FR01441.
 XX

PR 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 XX
 PI Bahr G, Cocude C, Capron A;
 XX
 DR WPI: 2002-082898/11.
 XX
 PT New polypeptide, useful for treating and diagnosing cancer or
 PT inflammation, and drug screening, comprises a human polynucleotide
 PT homologous to RNA helicase
 XX
 PS Example 2; Page 93-94; 114pp; French.
 XX
 CC The present invention relates to human RH16 (see AAA47798). RH16 is a
 CC 116kDa protein and has homology to RNA helicases (DEX box). RH16 and
 CC its coding sequence are useful for treating cancer; acute or chronic
 CC infections (especially by HIV or hepatitis B or C); inherited genetic
 CC diseases; (auto)immune diseases (particularly rheumatism, arthritis,
 CC arteriosclerosis, osteoporosis and diabetes, but many others listed),
 CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.
 XX
 SQ Sequence 1284 BP; 489 A; 198 C; 261 G; 335 T; 1 other:
 Alignment Scores:
 Pred. No.: 2,48e-155 Length: 1284
 Score: 1995.00 Matches: 378
 Percent Similarity: 99.47% Conservative: 0
 Best Local Similarity: 99.47% Mismatches: 2
 Query Match: 37.56% Indels: 0
 DB: Gaps: 0
 US-09-515-363c-2 (1-1025) x ABA04913 (1-1284)
 QY 646 AspGluGlyGlyAspAspGlyTyrCysAspGlyLysAspGluAspGluAspLeuLys 665
 Db 2 GATGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 61
 QY 666 ProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhePheLeuAsnLys 685
 Db 62 CTTTGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
 QY 686 MetLeuLysArgLeuAlaGluAsnProGlyTyrGluAsnGluLysLeuThrLysLeuArg 705
 Db 122 ATGTGAAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181
 QY 706 AspThrIleMetGluGlnIleTyrThrArgThrGlnGluSerPheAlaArgGlyIleLeuPheThr 725
 Db 182 AATACCATATGAGCAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 QY 726 LysThrArgGlnSerAlaTyrAlaLeuSerGlnTyrPheThrGluAsnLysPheAla 745
 Db 242 AAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 301
 QY 746 GluValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGluPheLysTyr 765
 Db 302 GAAGTGGAGTCAAAACCCACCATCTGATTTGGAGCTGACACAGCAGTGAAGTTCAAAACC 361
 QY 766 MetThrGlnAsnGluGlnLysGlyValIleSerLysPheArgThrGlyLysLysLeu 785
 Db 362 ATGACACAGAAATGAACAAAAGAGCATATGAAATTTTGGACCTGGAAATAAATTCG 421
 QY 786 LeuIleAlaThrThrValAlaGluGluGlyLeuAspIleLysglucysasnIleValIle 805
 Db 422 CTTATGCTACACACAGTGGCAGAGAGCTGAGATATTAAGAAATGTAACATTTTATC 481
 QY 806 ArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArgGlyValArgAla 825
 Db 482 CGTTATGCTGCTGCTCACCATGAAATAGCCATGCTCAGAGCCCGTGTGACCCAGACGCT 541

| | | | |
|--|----|--|-----------------|
| | PR | 20-OCT-2000; | 2000US-0241785. |
| | PR | 20-OCT-2000; | 2000US-0241786. |
| | PR | 20-OCT-2000; | 2000US-0241787. |
| | PR | 20-OCT-2000; | 2000US-0241808. |
| | PR | 20-OCT-2000; | 2000US-0241809. |
| | PR | 20-OCT-2000; | 2000US-0241826. |
| | PR | 01-NOV-2000; | 2000US-0244617. |
| | PR | 08-NOV-2000; | 2000US-0246474. |
| | PR | 08-NOV-2000; | 2000US-0246475. |
| | PR | 08-NOV-2000; | 2000US-0246476. |
| | PR | 08-NOV-2000; | 2000US-0246477. |
| | PR | 08-NOV-2000; | 2000US-0246478. |
| | PR | 08-NOV-2000; | 2000US-0246523. |
| | PR | 08-NOV-2000; | 2000US-0246524. |
| | PR | 08-NOV-2000; | 2000US-0246525. |
| | PR | 08-NOV-2000; | 2000US-0246526. |
| | PR | 08-NOV-2000; | 2000US-0246527. |
| | PR | 08-NOV-2000; | 2000US-0246528. |
| | PR | 08-NOV-2000; | 2000US-0246532. |
| | PR | 08-NOV-2000; | 2000US-0246609. |
| | PR | 08-NOV-2000; | 2000US-0246610. |
| | PR | 08-NOV-2000; | 2000US-0246611. |
| | PR | 08-NOV-2000; | 2000US-0246613. |
| | PR | 17-NOV-2000; | 2000US-0249207. |
| | PR | 17-NOV-2000; | 2000US-0249208. |
| | PR | 17-NOV-2000; | 2000US-0249209. |
| | PR | 17-NOV-2000; | 2000US-0249210. |
| | PR | 17-NOV-2000; | 2000US-0249211. |
| | PR | 17-NOV-2000; | 2000US-0249212. |
| | PR | 17-NOV-2000; | 2000US-0249213. |
| | PR | 17-NOV-2000; | 2000US-0249214. |
| | PR | 17-NOV-2000; | 2000US-0249215. |
| | PR | 17-NOV-2000; | 2000US-0249216. |
| | PR | 17-NOV-2000; | 2000US-0249217. |
| | PR | 17-NOV-2000; | 2000US-0249218. |
| | PR | 17-NOV-2000; | 2000US-0249244. |
| | PR | 17-NOV-2000; | 2000US-0249245. |
| | PR | 17-NOV-2000; | 2000US-0249284. |
| | PR | 17-NOV-2000; | 2000US-0249285. |
| | PR | 17-NOV-2000; | 2000US-0249297. |
| | PR | 17-NOV-2000; | 2000US-0249299. |
| | PR | 17-NOV-2000; | 2000US-0249300. |
| | PR | 01-DEC-2000; | 2000US-0250160. |
| | PR | 01-DEC-2000; | 2000US-0250391. |
| | PR | 05-DEC-2000; | 2000US-0251030. |
| | PR | 05-DEC-2000; | 2000US-0251988. |
| | PR | 05-DEC-2000; | 2000US-0256719. |
| | PR | 06-DEC-2000; | 2000US-0251479. |
| | PR | 08-DEC-2000; | 2000US-0251856. |
| | PR | 08-DEC-2000; | 2000US-0251868. |
| | PR | 08-DEC-2000; | 2000US-0251869. |
| | PR | 08-DEC-2000; | 2000US-0251989. |
| | PR | 08-DEC-2000; | 2000US-0251990. |
| | PR | 11-DEC-2000; | 2000US-0254097. |
| | PR | 05-JAN-2001; | 2001US-0259678. |
| | XX | (HUMA-) HUMAN GENOME SCI INC. | |
| | PA | Rosen CA, Barash SC, Ruben SM; | |
| | Pt | WIPI: 2001-465566/50. | |
| | DR | P-PsDB: AAU23647. | |
| | XX | | |
| | PT | Novel polypeptides and polynucleotides useful for diagnosing, | |
| | PT | preventing, treating neural, immune system, muscular, reproductive, | |
| | PT | pulmonary, cardiovascular, renal, proliferative disorders and cancerous | |
| | PT | diseases - | |
| | XX | | |
| | PS | Claim 4; SEQ ID NO 743; 1180bp; English. | |
| | CC | The present invention relates to the isolation of novel human enzyme | |
| | CC | polypeptides (AAU22915-AAU23614), and the cDNA and genomic sequences | |
| | CC | encoding them. The enzyme polypeptides of the invention may comprise the | |

CC functional classes of oxidoreductases, transferases, hydrolases, lyases
CC isomerase or ligase. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC **AA540785-AA541664** represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

50 Sequence 1319 BP; 491 A; 217 C; 294 G; 314 T; 3 other;

| | |
|------------------------|----------|
| Alignment Scores: | |
| Pred. No.: | 1,5e-152 |
| Score: | 1961.50 |
| Percent Similarity: | 93.64% |
| Best Local Similarity: | 92.27% |
| Query Match: | 36,93% |
| DB: | 22 |
| | |
| Length: | 1319 |
| Matches: | 406 |
| Conservative: | 6 |
| Mismatches: | 17 |
| Indels: | 16 |
| Gaps: | 4 |

US-09-515-363C-2 (1-1025) x AAS41517 (1-1319)

QY 442 ILeaspGLucYshIshIthrAsnLysGluAlaValTYrAsnAsnIleMetArgHisTYr 461

Db 23 GTTGATGATGTCATCAGACCAACCAAGAGCAGTGTATAATAACATCATGAGGCAT 82

| | | | |
|----|-----|---|-----|
| QY | 462 | LeumetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProVal | 481 |
|----|-----|---|-----|

DB 83 TTGATGCAGAGTTGAAAACATAGACTCAGAGAGAAACAAACCAGTGAATCCCTT 142

482 ProGlnIleuGlyLeuIhrIAsaEProGlyValGlyGlyAlaIhrLysGlnAlaIys 501

UD 143 CCICAGALNC1GGGAC1AACAGC11CAC1GG1G11GGAGGGCCACAGAGGCAGGC1.AAA 202

[illegible][illegible]

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858

On 563 C] nptb-tturyCvscJ nMetSsrPrMotSsrAenDhacJ wthrcJ nDrcTuvrcJ wJ wJ ITn 581

Db 383 CAACTTATTGTCAAATGAGTCCATGTCCAGTTTGGAAGTCAACCTATGAAACAATGG 442

582 AlaIleGlnMetGluValValAlaAlaIleValValGluValAsnArgIleValGluArgValCysAla 601

Db 443 GCCATTCAATGCAAAAAGCTGCAMAGAGGAATCGCAAGACGTGTTGTGCA 502

QY 602 GLHISLEUARQLSTYRASNGIUAAlaLeuGlnIleASNAsPThrIleArqMetIleAsp 621

Db 503 GAACATTGAGGACTACAAATGAGGCCCTACAAATTAATGACACCAATTGGAATGATATCAT 562

Qy 622 AlatyThrHisLeuGlutThrphetyrAsnGlulLysaspLysLysPheA:avai1le 641

Db 563 GCGTATACTCATCTTGAACCTTCTATATGAAGAGAAGATAGAAGTTGCAAGCATA 622

QY 642 GIuAspAspSerAspGluGlyGlyAspAspGluTyrCysAspGlyAspGluAspGluAsp 661

Db 623 GAAGATGATAGTGATGAGGGTGGTGATGATGAGTATTGTGATGATGGTGATGAACATGAGGAT 682

OY 662 AspleuLysProleuLysLeuaspGluThrAspArgPheLeuMetThrLeuPhe 681
 DB 683 GATTAAACAACTTGAACCTGGATGTAACAGATGATTTCTCATGATTTATTTT 742
 OY 682 GluAsnAspLysMetLeuLysArgLeuAlaGluAsnProGlyThrGluAsnGluLysLeu 701
 DB 743 GAAACAAATTAATTTGTTGAAAGGCTGGCTGAAACCCCAATATGTAATGAAAGCTG 802
 OY 702 ThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgThrGluGluAsnAlaArgGly 721
 DB 803 ACCAAATTAAGAAATACCAATATGAGCAATATAGAGATGAGCAATATACACAGAGCA 862
 OY 722 IleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnThrIleThrGluAsn 741
 DB 863 ATTAATCTTACAAAACACAGACAGATGATGCTTCCCTGAGTACTGAGATTAAGTAAAT 922
 OY 742 GluLysPheAlaGluValGlyValLysAlaHisLeuIleGlyValAlaHisSerSer 761
 DB 923 GAAATTTGCTGAGCTAGAGCTGCAAGCCCACTCTATGAGCTGACACAGCT 982
 OY 762 GluPheLysProMetThrGlnAsnGluGlnLysGluValLysSerLysPheArgThrGly 781
 DB 983 GATTCAAAACCATGACACAGATGAAACAAAAGAGCTATAGTAAATTCGCACTGCA 1042
 OY 782 LysIleAsnLeuLeuIleAlaThrThrValAlaGluGluLysLeuAspIleLysGluLys 801
 DB 1043 AAATTAATCTGCTTATGCTCTCC-ACAGTGGCAGAAAGATTTCTGATATTAAGCATCT 1101
 OY 802 AsnIleValIleArgTyrGlyLeuValThrAsnGluIleValMetValGlnAlaArgGly 821
 DB 1102 AACATTGKATTCGTTATGCTCTGCTCCACCAATGAAATGCC-ATGCTCCAGCC-CTGGCT 1159
 OY 822 ArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlyValAlaIle 841
 DB 1160 CGAGTAGAGCTGATGAGACACCTACGCTGCTGCTCA--CAGTGGTACGAGTATAC 1216
 OY 842 GluHisGluThrValAsnAspPheArgGluLysMetMetLysAlaIleHisCysVal 861
 DB 1217 GACTT-GAACAAGCTA--TGATTTCGAGAGAAA-----TGATCTATA 1254
 OY 862 GluAsnMetLys-----ProGluGluTyrAlaHisLysIleLeuGlnLeuGlnMet 878
 DB 1255 AGTATCATGCGTCAATATACACAGAGAGTT-GCTTATACAAATTTGGATTACAGATC 1313
 RESULT 10
 AAH15201
 ID AAH15201 standard; cDNA: 2613 BP.
 AC AAH15201:
 DT 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:13298.
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 PF 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 PS
 XX Claim 8: SEQ ID 13298; 2537bp + CD ROM; English.
 CC
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 3'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92445 to
 CC AAH3893 represent human amino acid sequences; and AAH1629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2613 BP; 591 A; 737 C; 783 G; 502 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,86e-101 Length: 2613
 Score: 1344.00 Matches: 291
 Percent Similarity: 59.55% Conservative: 136
 Best Local Similarity: 40.59% Mismatches: 238
 Query Match: 25.31% Indels: 52
 DB: Gaps: 12
 US-09-515-363C-2 (1-1025) x AAH15201 (1-2613)
 OY 306 LeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIle 325
 DB 224 ATGAGCTTGGTCTCTCAACAAAGGAGGTGATCATGCTGCTGAGAGGCAAAATATC 283
 OY 326 IleIleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrTleAlaLysAsp 345
 DB 284 ATCATCTGGCTGCCAGCGGTGCCGGAAGACCGCGCGTGTATAGTGGCAACGGG 343
 OY 346 HisLeuAspLysLysLysLysAlaSerGluProGlyLysValIleValLeuValLysLys 365
 DB 344 CACCTAGAG-----ACTGTGATGAGAGCAAGGTGTGATTTGTTATGCAACAGG 391
 OY 366 ValLeuLeuValGluGlnLeuPheArgLysGlnPheGlnProPheLeuLysLysTyrTrp 385
 DB 392 GTGCACCTGGTACCAG--CATGGTGAACAGTTCCAGGCCCATGCTGATGATGATGG 448
 OY 386 ArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPheProGluValValLys 405
 DB 449 ACCGTGACACCCCTGAGTGGGACATGGACACCTGCTGCTGGCCACCTGATCCGCG 508
 OY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLeuAsnLeuGlu 425
 DB 509 TGCCATGACCTGCTCATCTGCACACAGAGCTTGTGCAAGATGACATGACACCGCGAG 568
 OY 426 AsnGlyLysAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLysP2Lys 445
 DB 569 -----GAGGAGAGACCGAGCTGAGCTGCTCTTCTCTGATGTCGTGGTATAGATGC 622

XX 08-SEP-2000: 2000MO-US24704.
 PF 08-SEP-1999: 99US-0152921.
 PR 20-OCT-1999: 99US-0160575.
 PR 20-JAN-2000: 2000US-0177104.
 PR 07-SEP-2000: 2000US-0656633.
 XX
 PA (CURA-) CUREGEN CORP.
 PA (BIOJ) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR P-PSDB: AA000286.
 XX
 PT New interferon induced polypeptides and polynucleotides, useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 PS
 PS Claim 9: Page 29-32; 134pp; English.
 CC The sequence represents interferon induced nucleic acid, IFN α , IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SQ Sequence 1258 BP; 399 A; 267 C; 276 G; 316 T; 0 other:
 Alignment Scores:
 Pred. No.: 8,56e-90 Length: 1258
 Score: 1201.50 Matches: 242
 Percent Similarity: 94.53% Conservative: 0
 Best Local Similarity: 94.53% Mismatches: 1
 Query Match: 22.62% Indels: 14
 DB: 22 Gaps: 1
 US-09-515-363c-2 (1-1025) x AAS01149 (1-1258)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20
 DB 179 AATGTCGAATGGGTATTCACAGACGAGAAATTCGCTATCTATCTCGTCTTCAGGGCC 238
 QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuProAla 40
 DB 239 AGCGTGAATAATGTACTCA----- 258
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu 60
 DB 259 GAGGTCAAGGACGACATTAGAGAGACAGTCCGACCTCCGGAGACATGATGCCAGTTGAA 318
 QY 61 LeuLeuLeuSerThrLeuGluLysGlyValTyrPheLysGlyTyrPheArgLysPheVal 80
 DB 319 CTGCTCTCAGCAGCCTTGAGAAAGGAGCTGCGACCTTGCTGGACTCGCAATTCGCG 378

QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProAlaLeuThr 100
 DB 379 GAGGCCCTCCGGAGAACCGGACGCTCTGGCCGCCCTACATGAACTTACGCTCAGC 438
 QY 101 AspleuProSerProSerPheGluAsnAlaHisAspGlyTyrLeuGlnLeuLysLeu 120
 DB 439 GACTTGCCCTCTCCATCGTTTGAGAACGCTCATGTGATGAAATCTCCAACTGCTTAACTCG 498
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGlu 140
 DB 499 CTTCAGCCCACTCTGTGTGACAGAGCTTACTAGACAGCTTGTGATTAAGTATGAGAG 558
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnAsnTyrLys 160
 DB 559 GAGGAACCTCTGACAAATGACAGACAGAAACCGGATTCGTGTCGAGAAACAAATGAAAT 618
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTyrPheSerAla 180
 DB 619 GATCAGGTGTAGAGAGAGCTACTAAAGGATGTGCGAGAAAGAAAGAACTGCTTCTCCA 678
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnLeuLeuThrGlySer 200
 DB 679 TTTCGAATGTTCTTCCTGCAACAGAAACAAATGAACTTGTCCAGAGCTTAACAGCTCT 738
 QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyTyrGlnVal 220
 DB 739 GATTCCTCAGAAAGCAATGACAGCATGGAATTATGCAAGTATGATGCTCTCAACG 798
 QY 221 GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValTyrPheLys 240
 DB 799 GAAAGCAACTCTTCTTCAACCACTTCAAGCCAAATCTGGAGAGAGAGCTCTGCGCATG 858
 QY 241 GluAsnAsnSerSerGluSerSerPheAlaAspSerSerValValSer 256
 DB 859 GAGAAATACCTCATCAGAAATCATCTTTCACAGATCTTCTGTAGATTCA 906
 DB
 RESULT 12
 AAS01151
 ID AAS01151 standard; DNA; 1270 BP.
 XX
 AC AAS01151:
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Interferon induced nucleic acid, IFN α .
 XX
 DE Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KW haematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN α ; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX CDS 1..927
 XX FT /*tag= a
 XX FT /product= "IFN α "
 XX FT /note= "Interferon induced polypeptide"
 XX
 XX MO200118208-A2.
 XX
 XX PD 15-MAR-2001.
 XX
 XX PF 08-SEP-2000; 2000MO-US24704.
 XX
 XX PR 08-SEP-1999; 99US-0152921.
 XX PR 20-OCT-1999; 99US-0160575.
 XX PR 20-JAN-2000; 2000US-0177104.
 XX PR 07-SEP-2000; 2000US-0656633.

XX (CUBA-) CUBAGEN CORP.
 PA (BIO) BIOGEN INC.
 XX
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 XX MPI: 2001-235201/24.
 DR P-PSDB: AAU00298.
 XX
 PT New interferon induced polypeptides and polynucleotides, useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 9, Page 40-43; 134pp; English.
 XX
 CC The sequence represents interferon induced nucleic acids, IFN α , IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC hematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurotrophic diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SO Sequence 1270 BP; 401 A; 271 C; 279 G; 319 T; 0 other:
 Alignment Scores:
 Pred. No.: 8, 67e-90 Length: 1270
 Score: 1201.50 Matches: 242
 Percent Similarity: 94.53% Conservative: 0
 Best Local Similarity: 94.53% Mismatches: 1
 Query Match: 22.62% Indels: 14
 DB: 22 Gaps: 1
 US-09-515-363C-2 (1-1025) x AAS01151 (1-1270)
 OY 1 MetSerAsnGlyTyrSerThrAspGluAsnPhaArgTyrLeu|||SerCysPheArgAla 20
 DB 191 ATGTGCAATGGTATTCACAGACGAGAAATTCGGCTATCTTCTCGTCTTCACAGCC 250
 OY 21 ArgValysMetTyrIleGlnValGluProValLeuAspTyrIleThrPheLeuProVal 40
 DB 251 AGGCTGAATGTACATCA----- 270
 OY 41 GluValysGlnIleGlnArgThrValAlaThrSerGlyAsnMetGluAlaValGlu 60
 DB 271 GAGGTGAAGAGCAGATTCGAGACAGTGGCACCCTCCGAGAACATGACGGACAGTTGAA 330
 OY 61 LeuLeuLeuSerThrLeuGlnGlyValTyrPheIleGlnTyrThrArgIlePheVal 80
 DB 331 CTCTGCTGAGCACCCTGGAGAGGAGAGTCTGGACCTTGGTGGACTCCGCAATTCGTG 390
 OY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetSerGluLeuThr 100
 DB 391 GAGGCCCTCCGAGAACCGGACCCCTGGCCGCCCGCTACATGAACTTGGAGCTCAGC 450
 OY 101 AspLeuProSerProSerPheGluAsnAlaHisAspIleTyrLeuGlnIleLeuAsnIle 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 451 GACCTGCCCTCCATCGTTTGAGAACGCTCATGATGATATCTCCAACTGTTTAACTTC 510
 OY 121 LeuGlnProThrLeuValAspIleLeuValArgAspValLeuAspIleSerValu 140
 DB 511 CTTCAGCCCACTGTGTGGACAAAGCTTCTAGTAGAGAGCTTGGATGAGTGGATGAG 570
 OY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAlaAsnAsnGlyAsn 160
 DB 571 GAGGAACCTTGTACATTCGTAAGACAGAAACCGCATGCTGCTGGCGAAACATTCGAAT 630
 OY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTyrPheSerAla 180
 DB 631 GAATCAGGTGTAAAGAGACTACTAAAAAGATTGTGCAGAAAGAAACTGGTCTCTCA 690
 OY 181 PheLeuAsnValLeuArgIleThrGlyAsnAsnGluValGlnGluLeuThrIleSer 200
 DB 691 TTTCGATGATGTTCTTGCTTAACAGAGAAACAAATGAACTTGCCAAAGCTTAACAAGTCT 750
 OY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyTyrGlnVal 220
 DB 751 GATGCTCGAAGAACCATGACAGATTCGATGATTTATCACAAGTTGATGCTTCACTG 810
 OY 221 GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValTyrPheMet 240
 DB 811 GAAGAGCAACTCTTTCACACACAGTTCAGCCAAATCTGGAGAGAGGTCTGCTGATG 870
 OY 241 GluAsnAsnSerSerGluSerPheAlaAspSerValValSer 250
 DB 871 GAGAAATCACTACAGAAATCACTTTTGACGATTCCTGTGATGTTCA 910
 RESULT 13
 AAS91688
 ID AAS91688 standard; cDNA; 956 BP.
 AC AAS91688;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #27492.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX W0200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-0508631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR MPI: 2001-639362/73.
 DR P-PSDB: ABG27501.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 1, SEQ ID No 27492; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (1) is useful in gene therapy techniques
 CC to restore normal activity of (11) or to treat disease states involving
 CC (11). (11) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (11) and its binding partners are useful in medical
 CC imaging of sites expressing (11). (11) and (11) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WJPO
 CC at ftp.wjpo.int/pub/published_pct_sequences.

XX SQ Sequence 956 BP; 265 A; 236 C; 243 G; 212 T; 0 other;

Alignment Scores:

Pred. No.: 4,18e-82 Length: 956
 Score: 1106.50 Matches: 237
 Percent Similarity: 91.19% Conservative: 1
 Best Local Similarity: 90.80% Mismatches: 5
 Query Match: 20.83% Indels: 19
 DB: 23 Gaps: 1

US-09-515-363c-2 (1-1025) x AAS91688 (1-956)

QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPhenArgTyrLeuLeuSerGlyPheArgAla 20
 Db 191 ATGTCAATGGGTATTCACACAGCAAGATTCGCTACTCTCTGCTGCTTCACAGGCT 250
 QY 21 ArgValIleMetTyrIleGlnValGluProValIleuAspTyrLeuThrPheLeuProVal 40
 Db 251 AGGCTAAATGATGACATCCA----- 270
 QY 41 GluValIleGlnGlnIleGlnArgThrValAlaIleSerGlyAsnMetGlnAlaValGlu 60
 Db 271 GAGGTAAAGAGAGCATTCAGAGACAGTCCGACCTCCGGCAACATGCACACAGTGA 330
 QY 61 LeuLeuLeuSerThrLeuGluIleGlyValTyrPheIleGlyTyrPheArgGluPheVal 80
 Db 331 CTGCTGCTGAGCAGCTTGAGAAAGGAGTCTGACACCTTGCTGGCTCCCAATTGCTG 390
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 Db 391 GAGGCCCTCCGAGAACCGGACCCCTGCGCCGCGCTACATGAACTGAGCTGACG 450
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnGluLeuAsnLeu 120
 Db 451 GACTTCCCTCTCCATGCTTGGAAGAGCTCATGATGAATCTCCACCTGCTGAACT 510
 QY 120 uleuGln-ProThrLeu-ValAspIleLeuValArgAspValLeuAspIleCysMet 139
 Db 511 CCTTGAAGCCCACTCTGGGGGAGCAAGCTCTAGTAAAGCTGCGAI TAAGTCAAG 570
 QY 140 GluGluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaIleAsnAsnGly 159
 Db 571 GAGGAGAACCTGTTGACATTCAGAACAGAACCGGATTCGCTCCACAAAACATGGA 630
 QY 160 AsnGluSerGlyValArgGluLeuLeuLysArgIleValGlnGluLysAsn-Triphe 179
 Db 631 AATGAATCAAGGTAAAGAGAGCTAAAGAAAGAGATTGCTCAGAAACAAACCTGCTTC 690
 QY 179 IAlaPheLeuAsnValIleuArgGlnThrIleLysAsn-AsnGluLeuValGlnGluLeuThg 199
 Db 691 TGCATTTCGAATGTTCTTCCTGCAAAACAGAAACCAATGAAGTTCGCAAGATTAAAG 750
 QY 199 IYserAspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyProG 219
 Db 751 GCTCTGATTGCTCAGAAACCAATGACAGATTGAGAAATTATTCACAAAGTTAATGGCTC 810

QY 219 IValGluGlnIleuLeuSerThrThrValGlnProAsnLeuGluIleGluValTrig 239
 Db 811 AAGTGAAGAGCAAGCTTCTTCAACCAAGTTCACAGCAAAATCTGAGAAAGCAAGTTCAG 870
 QY 239 IYMetGluAsnSerSerGluSerSerPheAlaAspSerSerValIleSer 256
 Db 871 GCATGAGACATTAACATCATCAGATCATCTTTGACAGATTCTTCTACTTCA 273
 RESULT 14
 AAS40969
 ID AAS40969 standard; cDNA, 1978 BP.
 AC AAS40969;
 DT 17-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human enzyme polypeptide #185.
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 OS Homo sapiens.
 XX
 PN WO20015301-A2.
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214866.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220964.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.

| | | | |
|----|------|---|------|
| QY | 500 | AlaLysAlaGluGlnHisIleLeuLysLeuGlyAlaAsnLeuSfPalThrThrlLys | 513 |
| Db | 111 | GATGGGCGCATCAACACGCTCGACGCTTGCGCAACTTCTACAGCTGGTCATG | 170 |
| QY | 520 | ThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnIleProValLysIle | 539 |
| Db | 171 | TCACCCGAGAACTGCTGCCCCAGAGCTGCAGAGACACAGCAACAGCTTGTAAACAGTAC | 230 |
| QY | 540 | AlaIleAlaAspAlaThrArgLysAspProPheLysGluLysLeuLeuGlnIleThr | 559 |
| Db | 231 | AACCTGCGCACAGCGCCACGACGAGATCCGTTGGGAGCTTGTACAGAACCTCATAC | 290 |
| QY | 560 | ArgIleGlnThrTrpCysGlnMetSerPrometSer---AspPheThrGlnProTrp | 578 |
| Db | 291 | CAATTCATGACCCACCTCGGAGATGCTGACTGATGACCGGAATTTGGACGCAATGTAT | 350 |
| QY | 579 | GluGlnIleTrpAlaIleGlnMetGluLysLysAlaAlaLysLysGlyAsnAlaProGluArg | 598 |
| Db | 351 | GAGCAGACGAGTGGTGAAGCTGAGTACGAGCGCGCTTGCTGCTTCACAGCAACG | 410 |
| QY | 599 | ValLysAlaGluHisLeuArgLysTrpAsnGluAlaLeuGlnIleAsnAspThrIleArg | 618 |
| Db | 411 | GTGTATCGGCTTACCTCGAGCGGCTCAATGACGCGCTCATCATCAACCTCGCGC | 470 |
| QY | 619 | MetIleAspAlaTrpThrHisLeuGluThrPheTrpAsnGluGlnLysAspLysIlePhe | 638 |
| Db | 471 | GCGGTGATGCTGCGTGGCGCTGCAGAGATTTCATACAGGACGACAGCTACTAAACG | 530 |
| QY | 639 | AlaValIleGluLysAspSerAspGluGlyLysAspAspLysTrpCysAspGlyAspLeu | 658 |
| Db | 531 | CAGATCTCGT-----TGT----- | 542 |
| QY | 659 | AspLysAspAspLeuLysLysProLeuLysLeuAspLysIleAspArgPheLeuMetThr | 678 |
| Db | 543 | -----GCTTAAGCGCCACTGCTTACGC | 563 |
| QY | 679 | LeuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGlnTrpGluAsn | 698 |
| Db | 564 | CTGTTCATGATACCGACAGATGAGCTGGCCACTTGCAACCTATGCGCTA---GAGAT | 620 |
| QY | 699 | GluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTrpTrpArgThrGluGlnSer | 718 |
| Db | 621 | CCAAACCTGGAGATGCTGGAAAGATCTCGCAAAAGCACCTTGAAGT---ATTCTAACAC | 677 |
| QY | 719 | AlaArgGlyIleIlePheThrLysThrArgGlnSerAlaTrpAlaLeuSerGlnIleProIle | 738 |
| Db | 678 | CTGTGCGGTATCATCTTCAACCCGACCCGCCAAAGGCGCACTCCCTCTACTCTGGCTC | 737 |
| QY | 739 | ThrGluAsnGluLysPheAlaGluValGlyLysAlaHisHisLeuIleGlyAlaCys | 758 |
| Db | 738 | CACGACGACAGAGGCGCTGCAGACTGTGGAGATCCGAGCGGCAGCTACTGATGGGCTGG | 797 |
| QY | 759 | HisSerSerGluPheLysPrometThrGlnAsnGluGlnGluValIleSerLysIle | 778 |
| Db | 798 | AACGCGACCGACAGACACCCATGACCCGACAG-GACCACTAGAGATGCTACGAACTTC | 856 |
| QY | 779 | ArgThrGlyLysIleAsnLeuLeuIleAlaThrThrValAlaGluGluLysAspIle | 798 |
| Db | 857 | CAACATGGAACCTGAACCTTCTGTGTGGCCACAGAGTGTGGGTAGAGAGAGCTGCACATC | 916 |
| QY | 799 | LysGluCysAsnIleValIleArgTrpGlyLeuValThrAsnGluIleAlaMetValGln | 818 |
| Db | 917 | CCACATTGCAATGTGTGTGGCTATATGGGCTCTTGACCAATGAATGTATGTATGGTACG | 976 |
| QY | 819 | AlaArgGlyArgAlaArgAlaAspGlnSerThrTrpValLeuValAlaHisSerGlySer | 838 |
| Db | 977 | GCCAGGGGCGCGTCCCGGCGCATGAGATGTAACGCTTGTTGTGCAATGTGAAGTGTC | 1036 |
| QY | 839 | GlyValIleGluHisGluThrValaAsnAspPheArgGluLysMetMetTrpLysAlaIle | 858 |
| Db | 1037 | CGGAGCTGAAGCGGAGCTGATCAACGAGCGCCTGGAGACGCTCATGTGACATGAGCTGTG | 1096 |

| | | | | |
|-----------|---|--------|--|------|
| OY | | 859 | HlCysValGlnAsnMetLysProGluGluTyrAlaHisLysLeuIleuValThrIleMet | 878 |
| Dd | | 1097 | GCTGCTGTGCAGAAATATGGACCAAGCCACTACCAGGCCAAGATCCGGATCATTTACAG | 1156 |
| OY | | 879 | GInserIleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsn | 898 |
| Dd | | 1157 | GCAGCGTGACCAACGGGGGGCCAGAGCACGCCACGGGGAGAACACAGCGCAATTC | 1216 |
| OY | | 899 | AsnProSerLeuIleThrPheLeuCysLysAsnCysSerValLeuAlaCysSerAllyLeu | 918 |
| Dd | | 1217 | CCAGTGGAGCAGCGAGCTACTCTGTCATCACTGCATGGTGTGCTGTGGGCTTAAGC | 1276 |
| OY | | 919 | AspIleHisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLeu | 938 |
| Dd | | 1277 | GACCTGGGAAGAAGGAGGAGCCACCACCATGTCAATGTACAACCCCAACTTCGAACTAC | 1336 |
| OY | | 939 | Tyr---IleValArgGluAsnLysAlaLeuGlnLysCysAlaAspTyrGlnIleAsn | 957 |
| Dd | | 1337 | TATATGTCTCCAGCGATCTCTGTGTCATCAACAAAGTCTTCAAGACATCGAAGGCTGAG | 1396 |
| OY | | 958 | GlyGluIlelleCysLys---CysGlyGlnAlaITripGlyThrMetMetValHisLysGly | 976 |
| Dd | | 1397 | GGTGTCTACACTCAGCTCCAGAACTGTGGGAGAGCTGGGGGTCTGAGATGATGTACAAAGCA | 1456 |
| OY | | 977 | LeuAspLeuProCysLeuLysIleArgAsnPheValValValPheLysAsnAsnThr | 996 |
| Dd | | 1457 | GTGAAGCTGGCCAGTCTCAAAGATCCGAGC-----ATGCTGTGGAGACCTTTACAGG | 1510 |
| OY | | 997 | LysLysGlnTyrLysLysTyrValGluLeuProIleThrPheProAsnLeuAsfTyrSer | 1016 |
| Dd | | 1511 | CGGATCCAGGCCAAAAGATGTGTCGGGGTCCCTTCCTTCGGTGGCTGACTTGAATTTCG | 1570 |
| OY | | 1017 | GluCysCys | 1019 |
| Dd | | 1571 | CAGCATTTGT | 1579 |
| RESULT 15 | | | | |
| ID | ASO1150 | | | |
| XX | ASO1150 standard; DNA; 3692 BP. | | | |
| AC | ASO1150; | | | |
| XX | | | | |
| DT | 12-SEP-2001 (first entry) | | | |
| XX | | | | |
| DE | Interferon induced nucleic acid, IFN5. | | | |
| XX | | | | |
| KW | Interferon induced nucleic acid; autoimmune disease; lupus erythematosus; | | | |
| KW | immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS; | | | |
| KW | graft rejection; viral infection; hepatitis; aplastic anemia; cancer; | | | |
| KW | human immunodeficiency virus; HIV; immune-mediated glomerulonephritis; | | | |
| KW | haematologic disease; chronic neutropenia; myocardial infarction; | | | |
| KW | neurological disease; Alzheimer's disease; Parkinson's disease; tumour; | | | |
| KW | amphotropic lateral sclerosis; spinal muscular atrophy; human; IFN5; ds. | | | |
| XX | | | | |
| SS | Homo sapiens. | | | |
| XX | | | | |
| Key | Location/Qualifiers | | | |
| FT | CDS | 1..804 | | |
| FT | /tag-a | | | |
| FT | /product-"IFN5" | | | |
| FT | /note-"interferon induced polypeptide" | | | |
| XX | | | | |
| PN | MO200118208-A2. | | | |
| PD | | | | |
| XX | 15-MAR-2001. | | | |
| Pf | 08-SEP-2000; 200OWO-US24704. | | | |
| XX | | | | |
| PR | 08-SEP-1999; 99US-0152921. | | | |
| PR | 20-OCT-1999; 99US-0160575. | | | |
| PR | 20-JAN-2000; 2000US-0177104. | | | |
| PR | 07-SEP-2000; 2000US-0656633. | | | |
| XX | | | | |

PA (CURA-) CUBAGEN CORP.
 PA (BIO) BIOGEN INC.
 XX
 PT Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR MPI: 2001-235201/24.
 DR P-PSDB: AAU00297.
 XX
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 9; Page 33-39; 134pp; English.
 XX
 CC The sequence represents interferon induced nucleic acid, IFN γ , IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC hematologic diseases such as aplastic anemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SQ Sequence 3692 BP; 1166 A; 709 C; 806 G; 1011 T; 0 other:
 Alignment Scores:
 Pred. No.: 5,32e-70 Length: 3692
 Score: 969.50 Matches: 196
 Percent Similarity: 91.67% Conservative: 2
 Best Local Similarity: 90.74% Mismatches: 5
 Query Match: 18,25% Indels: 14
 DB: 22 Gaps: 1
 US-09-515-363c-2 (1-1025) x AAS01150 (1-3692)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerC.sphenagAla 20
 DB 179 ATGCGAATGGGTATTCACAGACGAGAAATTCGCGATCATCTCGTATTCACGGCC 238
 QY 21 ArgValIysMetTyrIleGlnValAlaGluProValLeuAspTyrLeuThrPheLeuProAla 40
 DB 239 AGGGTGAATAATGTCATCCA----- 258
 QY 41 GluValIysGlnGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu 60
 DB 259 GAGGTGAAGGAGGATTCAGAGACAGTCGCCACCTCCGGGAACATGACAGCGATTGAA 318
 QY 61 LeuLeuLeuSerThrLeuGlnIulysGlyValIlePheIleGlnTyrThrArgGluPheVal 80
 DB 319 CTGCTGCTGAGCACCTGGAGAAAGGAGTGGACCTTGCTGGACTCCGGAATTGCTG 378
 QY 81 GluAlaLeuArgThrGlySerProLeuAlaIleArgTyrMetAsnProGluLeuThr 100
 DB 379 GAGGCCCTCCGAGAACCGGAGCCCTCTGGCGCCCGCTACATGAAACCTTGAAGCTACG 438
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnIleLeuAsnLeu 120
 DB 439 GACTTGCCCTCTCATCGTTTGAGAACGCTCATGATGATATATCTCCAAACTGCTGAACCTTC 498

QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysSerGlu 140
 DB 499 CTTCAGCCCACTCTGTGGACAAAGCTTCTAGTAGACGTTTGATTAAGTCATGAGAG 558
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaAlaGluAsnAsnGlyAsn 160
 DB 559 GAGGAACCTGTTCACATTTAAACAGAAACCGGATTGCTGCTGGCAAAACATATGCAAT 618
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpPheSerAla 180
 DB 619 GAATCAGGTGTAAAGAGAGCTACTAAAGAGATTGTGCAGAAAGAAACTGCTCTCTGCA 678
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrLysSer 200
 DB 679 TTTCGAATGTTCTTCTGTCAAACAGAAACAAATGAACTTGTCCAAAGATTAAACAGCT 738
 QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAsp 216
 DB 739 GATTGCTGAAAGCAATGCAGTATTTGTAATTTTACTGAGAGAGAT 784

Search completed: May 16, 2003, 17:37:09
 Job time : 432 secs

GenCore version 5.1.1.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 16:36:02 ; Search time 4480 seconds

(without alignments)
6658.566 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311

Sequence: 1 MSNGYSTDENFRYLISCFRA.....LPITFPNUDYSENCLEFSDSD 1025

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+2pn.model -DEV=xlh
-O=/gen2.1/USPTO/spool/US09515363/runat_16052003_100246_24816/app.query.fasta.1.1223
-DB=GenEmbl -QFRT=fastap -SUFF=trge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=Pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=Pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09515363 -CGCN_1.1.2966 -runat_16052003_100246_24816 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOC -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

1: GenEmbl:*
2: gb_da:*
3: gb_hg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_or:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_iny:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------------|---------------------|
| 1 | 5311 | 100.0 | 3380 | 9 AF095844 | AF095844 Homo sapi |
| 2 | 5285 | 99.5 | 3372 | 6 AX300832 | AX300832 Sequence |
| 3 | 5285 | 99.5 | 3373 | 9 AY017378 | AY017378 Homo sapi |
| 4 | 4199 | 79.1 | 3078 | 10 AY075132 | AY075132 Homo sapi |
| 5 | 4194 | 79.0 | 3771 | 10 AF374384 | AF374384 Mus muscu |
| 6 | 2874 | 54.1 | 1776 | 9 AK056293 | AK056293 Homo sapi |
| 7 | 2690.5 | 50.7 | 2468 | 10 BC025508 | BC025508 Mus muscu |
| 8 | 2442 | 46.0 | 1443 | 6 AX300841 | AX300841 Sequence |
| 9 | 2271.5 | 42.8 | 2050 | 10 BC004031 | BC004031 Mus muscu |
| 10 | 2021 | 38.1 | 1213 | 9 BC007966 | BC007966 Homo sapi |
| 11 | 1995 | 37.6 | 1284 | 6 AX300838 | AX300838 Sequence |
| 12 | 1356 | 25.5 | 2326 | 10 AF316999 | AF316999 Mus muscu |
| 13 | 1356 | 25.5 | 2336 | 10 BC029209 | BC029209 Mus muscu |
| 14 | 1352 | 25.5 | 2584 | 9 BC014949 | BC014949 Homo sapi |
| 15 | 1344 | 25.3 | 2613 | 9 AK021416 | AK021416 Homo sapi |
| 16 | 1201.5 | 22.6 | 1258 | 6 AX098232 | AX098232 Sequence |
| 17 | 1201.5 | 22.6 | 1270 | 6 AX098236 | AX098236 Sequence |
| 18 | 1129 | 21.3 | 2250 | 9 AK097669 | AK097669 Homo sapi |
| 19 | 1090.5 | 20.5 | 3065 | 9 AF038963 | AF038963 Homo sapi |
| 20 | 1023 | 19.3 | 163681 | 9 AC007750 | AC007750 Homo sapi |
| 21 | 1019.5 | 19.2 | 4024 | 4 AF181119 | AF181119 Sus scrofa |
| 22 | 969.5 | 18.3 | 3692 | 6 AX098234 | AX098234 Sequence |
| 23 | 969.5 | 18.3 | 3704 | 6 AX098238 | AX098238 Sequence |
| 24 | 783 | 14.7 | 95417 | 9 AC010876 | AC010876 Homo sapi |
| 25 | 769.5 | 14.5 | 210317 | 2 AC115074 | AC115074 Mus muscu |
| 26 | 708 | 13.3 | 3026 | 9 AK023661 | AK023661 Homo sapi |
| 27 | 687 | 12.9 | 3114 | 3 AF480439 | AF480439 Caenorhab |
| 28 | 678 | 12.8 | 923 | 10 BC019605 | BC019605 Mus muscu |
| 29 | 642 | 12.1 | 392 | 6 AX071769 | AX071769 Sequence |
| 30 | 636.5 | 12.0 | 182906 | 2 AC111919 | AC111919 Rattus no |
| 31 | 630 | 11.9 | 387 | 6 AX071783 | AX071783 Sequence |
| 32 | 616 | 11.6 | 3099 | 3 AF480440 | AF480440 Caenorhab |
| 33 | 532.5 | 10.0 | 118891 | 2 AC094719 | AC094719 Rattus no |
| 34 | 516 | 9.7 | 2633 | 9 HSM802366 | HSM802366 Pyrococu |
| 35 | 479 | 9.0 | 45931 | 3 US8757 | US8757 Caenorhabd |
| 36 | 479 | 9.0 | 226505 | 2 AC006899 | AC006899 Caenorhab |
| 37 | 472 | 8.9 | 301 | 6 I50897 | I50897 Sequence |
| 38 | 467 | 8.8 | 118891 | 2 AC094719 | AC094719 Rattus no |
| 39 | 457.5 | 8.6 | 182906 | 2 AC111919 | AC111919 Rattus no |
| 40 | 453 | 8.5 | 307150 | 1 CNSPAX01 | AJ248283 Pyrococu |
| 41 | 453 | 8.5 | 349980 | 6 AX041106 | AX041106 Sequence |
| 42 | 441.5 | 8.3 | 253505 | 1 AP000007 | AP000007 Pyrococu |
| 43 | 437 | 8.2 | 377 | 6 A74554 | A74554 Sequence |
| 44 | 437 | 8.2 | 377 | 6 A77533 | A77533 Sequence |
| 45 | 417 | 7.9 | 10782 | 1 AE001002 | AE001002 Archaeogl |

ALIGNMENTS

RESULT 1

AF095844
 LOCUS AF095844 3380 bp mRNA linear PRI 23-JAN-2002
 DEFINITION Homo sapiens melanoma differentiation associated protein-5 (MDA5)
 ACCESSION AF095844
 VERSION AF095844.1 GI:11344593
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3380)
 AUTHORS Kang, D.-C., Gopalakrishnan, R.V., Wu, Q., Jankowsky, T., Pyle, A.M. and
 Fisher, P.B.
 TITLE mda-5: An interferon-inducible putative RNA helicase with
 double-stranded RNA-dependent ATPase activity and melanoma
 growth-suppressive properties
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 637-642 (2002)
 MEDLINE 21664412
 PUBMED 11805321
 REFERENCES 2 (bases 1 to 3380)
 AUTHORS Kang, D.-C. and Fisher, P.B.
 TITLE Direct Submission
 JOURNAL Submitted (30-SEP-1998) Urology, Columbia University, P&S, 630
 W16th, New York, NY 10032, USA
 FEATURES
 source
 1..3380
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="HO-1"
 /issue_type="melanoma"
 1..3380
 /gene="MDA5"
 /gene="MDA5"
 169..3246
 /codon_start=1
 /product="melanoma differentiation associated protein-5"
 /protein_id="AAG33368.1"
 /db_xref="GI:11344594"
 /translation="MSNGYSTDENFRYLISCFRRKVKYIOVEPVLDYITFLPAEVE
 OIORVATSGMNAVEILLSTLEKGMILGMRREVEAIKRTGSLARYNPELTL
 PSPFENAHDELQNLLOPTLVOKILVROVLDKCMPELITIEDNRITAAENNG
 EGVRELLKRVOKENMSAFLNVIROGNNRIVQELTNSGSSNAETELNSQDQ
 OVEBGLSTTVQPNLEKVGWENNSFSRSDSVSSTLSAEGSVCLDSLGH
 NSNMGSDGTMGSDSDENVAARASPEPELI.KPQOMVAAPALEKNIITLPTGSG
 KTRVAVYIAKDLDKKKRSEPKVIVYANVLLVEQLKKEFQPLKMYRVLGSG
 DTQLISPEVAVKSDIITSTQILLENLNLINENEDAVOLDSPLIIDECHTNK
 EAVYNNIRHLYMOKLKNRLKKEKKPVIPDILIGLASGGAGATQAAAEHLIK
 LCANDARTITVKNLIDOLKNQIOEPRKRAIDATIRPKKLEIMTRIQTYCO
 MSPHSDFTQPTFOVAIDOMEKKAAGKRRKRVCAEHLKTKNEALQINDTRIMIDAT
 HLETTYNEEKQKRAVIDEDSDGDDCYGDIIDEDDIKKPLKIDEDRTIMLTFE
 NKKMLKRLAENPEYENKLTAKRLNIMQYRTIESAGIIFETKRGSAVLSQMITF
 NEKFAEVGKAHHILGAGHSEPKPTNBOKEPVISKPRCKINMLLATVAREGLDI
 KECNIVIRYGLTNEIANYOARGARABESTYVIVAAHSGVITHELTTPDREKMYK
 AHCYQNNKPEPTAKHILIELOMOSIMEKKTKRKIAKTKNNNSLTFTUCKNCSVLA
 CSGEDIHIEKMHVHNMTPERKELYIVENKALCKKAMVQINDEIICKCOAGNSTM
 VHKGLDLCLEIRNFVYFKNSSTKKYKWELEPITPNNIDYSECLFSDP"
 BASE COUNT 1153 a 644 c 753 g 830 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3380
 Score: 5311.00 Matches: 1025
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-515-363c-2 (1-1025) x AF095844 (1-3380)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerTyrPheArgVala 20
 DB 169 ATGTGCAATGGGATATTCACAGACGAAATTCCTCGCTATCATCTCGTCAAGGAGCC 228

QY 21 ArgValIysMetTyrIleGlnValGluProValIleuAspTyrLeuThrPheLeuIleVala 40
 DB 229 AGGTGAAATATGTACATCAGATGAGCGCTGTGCTGACACTACGTACCTTCTGCTGCA 288
 QY 41 GluValIysGluGlnIleGlnIleArgThrValAlaThrSerGlyAsnMetGlnAlaVala 60
 DB 289 GAGGTGAAGAGACAGATTCAGAGGACAGCTCCGACCTCGGGAAACATGACGCAATGAA 348
 QY 61 LeuLeuLeuSerThrLeuGlnIleGlyGlyValTrpHisLeuGlyTyrPheArgGluVala 80
 DB 349 CTGCTGCTGACGACCTTGGAGAGAGGAGCTGGACCTTGGTGGACTGGGAAATGTGTG 408
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluThr 100
 DB 409 GAGGCGCTCCGAGAACCGGACGCGCTGCGCGCGCGCTCATATGAACCTGAGATACG 468
 QY 101 AspleuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuVala 120
 DB 469 GACTTGCCCTCTCCATCTCTTGAGAACCTCATGTGATGATATCTCCAACTGCTGATCTC 528
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysSerGlu 140
 DB 529 CTTCAGCCCTACTGTGTGAGACAGCTTCTAGTAGAGCGTCTTGATAGCTGCTAGAG 588
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnAsnVala 160
 DB 589 GAGCAACTGTTGACAAATTCGAGAGACGAAACCGGATTCCTGCTGCAAGAAACAAAT 648
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpPheVala 180
 DB 649 GAATCAGGTGTAAAGACACTACTATAAAGATTTGCGAAGAAAGAAATGCTGTCTTCCA 708
 QY 181 PheLeuAsnValIleuArgIleThrGlyAsnAsnGluLeuValGlnGluLeuThrIleSer 200
 DB 709 TTTCGAATGTTCTTCTGCTCAACAGAGAAACAAATCACTTGCCAAAGATTAAATATCT 768
 QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValaAspGlyProIleVala 220
 DB 769 GATTCCTCAGAAAGCAATGACAGATTCAGATTATTCACAACTGATGATGCTCTCAATG 828
 QY 221 GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValaThrIleMet 240
 DB 829 GAAAGGCAACTCTTCTTACCAACACATTCAGCAAAATTCGAGAGAGAGGTTCGATGATG 888
 QY 241 GluAsnAsnSerSerGluSerSerPheAlaAspSerSerValaIleSerGluSerAspThr 260
 DB 889 GAGATTAATCTCATCATCATCTTCTTCCACATTTCTTGTAGTTTCAGATTCATATACA 948
 QY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMet 280
 DB 949 AGTTGGCAGAGAAAGAGTCAAGCTGAGCTTACATCAAAATCTTGGACATTAACATCAATG 1008
 QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluAsnValAlaAlaValaVala 300
 DB 1009 GCGCACTGATTCAGGACCATGAGGAGTATTCACATGAAGAGAAATGTGGCAATAGCA 1068
 QY 301 SerProGluProGluLeuGlnIleuArgProTyrGlnMetGluValAlaGlnIleProVala 320
 DB 1069 TCCCGGAGCCAGAACTCAGCTCAGGCTTACCAATAGAAAGTTGCCAGACATGACCTTG 1128
 QY 321 GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVala 340
 DB 1129 GAAGGAGAAATATCATCATCTGCTCCTACAGGAGATGAAAAACACAGACTGCTGTGT 1188
 QY 341 TyrIleAlaLysAspHisLeuAspLysLysLysAlaSerGluProGlyIleValIleCys 360
 DB 1189 TACATTGCCAAGATCACTTGACACAGAAAGAAAGCATCTGACCTCGCAAAATTTATA 1248
 QY 361 ValLeuValAsnLysValLeuLeuValGlnGlnLeuPheArgLysGluPheGlnIleGlnPhe 380
 DB 1249 GTTCTTGCATTAAGCTACTCTAGTTGACACAGCTCTTCCGCAAGAGGCTTCAATATT 1308

QY 381 LeuLysLysTrpPyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe 400
 DB 1309 TTGAGAAATGATGCTGTTATTGATTAAGTGGTGAATACCACTGAAATATATCATTT 1368
 QY 401 ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSer 420
 DB 1369 CCGAAGGTGTCAAGCTCTGTGATATTATATGATACAGCTCAAACTCTTGAAACTTC 1428
 QY 421 LeuLeuAsnLeuGlnAsnGlnGlyAspAlaGlyValGlnLeuSerAspIleSerLeuIle 440
 DB 1429 CTCTTAACTTGAATAATGAGAAAGATGCTGGTTCATTAATGCAACTTTCCTCATTT 1488
 QY 441 IleIleAspGluCysHisIleThrAsnLysGluAlaVal-TyrAsnAsnIleMetCysHis 460
 DB 1489 ATCATGATGATGATGATCATGACCAACAAAGAGAGTGTATATATACATCATGAGGCA 1548
 QY 461 TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGlnAsnLysProValIlePro 480
 DB 1549 TATTTGATGCGAAGTGTGAAACAAATAGACTCAAGAAAATAAACAAACTGATTCCT 1608
 QY 481 LeuProGlnIleLeuGlnLeuThrAlaSerProGlyValAlaGlyValAlaThrLysGlnAla 500
 DB 1609 CTTCCTCAGTACTGGCACTACACAGCTTCACTGCTGTGTGGAGGGCCAGAGCAAGCT 1668
 QY 501 LysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThr 520
 DB 1669 AAGGCTCAACACACATTTTAAACATATGCGCAATCTTATCATCATTTATTAATAACT 1728
 QY 521 ValLysGlnAsnLeuAspGlnLeuLysAsnGlnIleGlnLupProCysLysLysPheAla 540
 DB 1729 GTTAAACAAACCTTGTACTCACTGAAAAACCAATACAGCAAGCCATGCAAGAGTTGC 1788
 QY 541 IleAlaAspAlaThrArgGluAspProPheLysGlnLysLeuLeuGlnIleMetThrArg 560
 DB 1789 ATTGCAATGCAACACAGAGAGATCCATTAAAGAGAACTTCTAGAAATAAGACAAAG 1848
 QY 561 IleGlnThrTyrCysGlnMetSerPrometSerAspPheGlyThrGlnProTyrGlnAla 580
 DB 1849 ATTCAAACTTATGTCAAATGAGTCCAAATGTCAGATTTTGTAACTCAACCTGTAAACA 1908
 QY 581 TrpAlaIleGlnMetGlnLysLysAlaAlaLysLysGlnAsnArgLysGlnArgValCys 600
 DB 1909 TGGGCCATTCATAATGAAAAAAGCTGCAAAAAAAGAAATGCAAAAGCTGTTGTG 1968
 QY 601 AlaGlnLysLeuArgLysTyrAsnGlnAlaLeuGlnIleLeuAsnProThrIleArgMetIle 620
 DB 1969 GCGAACAATTTGAGAGAGTACAAATGAGGCCCTACAAATTAATGACACAAATGCAATGA 2028
 QY 621 AspAlaThrThrHisLeuGlnThrPheTyrAsnGlnLysAspLysLysPheAlaVal 640
 DB 2029 GATGCGTATACATCTTGAACCTTTCTTATATGAAAGAGAAACATTAAGAAGTTGCA 2088
 QY 641 IleGlnAspSerAspArgLysGlyLysAspAspLysTyrCysAspLysAspGln 660
 DB 2089 ATGAAAGATGATGTGATGAGGGGTGATGATGATGATGATGATGATGATGATGATGAT 2148
 QY 661 AspAspLeuLysLysProLeuLysLeuAspGlnThrAspArgPheLeuMetThrLeuPhe 680
 DB 2149 GATGATTTTAAAGAAACCTTTGAAACTGGAATGAAACAGATTTTCAAGACTTTAAT 2208
 QY 681 PheGlnAsnAsnLysMetLeuLysArgLeuAlaGlnAsnProGluTyrGlnAsnGlnLys 700
 DB 2209 TTTGAAAAACATATAATGTTGAAAAAGGCTGCTGAAAAACCCAGAAATATGAAATGAAA 2268
 QY 701 LeuThrLysLeuArgAsnThrIleMetGlnGlnTyrThrArgThrGlnGlnSerAlaArg 720
 DB 2269 CTGACCAAAATTAAGAAATACCAATATGACCAATATCTAGAGCTGAGGATGACAGCA 2328
 QY 721 GlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTrpIleThrGln 740
 DB 2329 GGAATATATCTTTACAAAAACAGACAGATGATGCTTCCCTCCAGCAATGATTAATGAA 2388
 QY 741 AsnGlnLysPheAlaGlnValAlaGlyValAlaHisIleLeuIleGlyAlaGlnLysSer 760

DB 2389 AATGAAAAATTTGCTGAAGTACAGAGTCAAAAGCCCAACATCTGATTGAGACTGACACAGC 2448
 QY 761 SerGluPheLysPrometThrGlnAsnGlnLysGlnValIleSerLysPheArgThr 780
 DB 2449 AGTGAGCTTCAAAACCCATGACACAGATGAAACAAAAAGAGTCAATTAGTAAATTCGCA 2508
 QY 781 GlyLysIleAsnLeuLeuIleAlaThrThrValAlaGlnGlnLysAspIleLysGln 800
 DB 2509 GGAATAATCAATCTGCTTATGCTTACACAGAGGCAAGAAAGGTCTGATATTAACAA 2568
 QY 801 CysAsnIleValIleArgTyrGlyLeuValThrAsnGlnIleAlaMetValGlnAlaArg 820
 DB 2569 TGTAACTTGTATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2628
 QY 821 GlyArgAlaArgAlaAspLysThrTyrValLeuValAlaHisSerGlySerGlyVal 840
 DB 2629 GGTGAGCCAGCAGCTGATGAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2688
 QY 841 IleGlnHisGlnThrValAsnAspPheArgGlnLysMetMetTyrLysAlaIleHisCys 860
 DB 2689 ATGCAACATGACACGATTTATGATTTCCGAGAGAGATGATGATTAAGCTATACATTT 2748
 QY 861 ValGlnAsnMetLysProGlnLysLysLysLysLysLysLysLysLysLysLysLys 880
 DB 2749 GTTCAAAATATGAAACACAGAGAGTATGCTCATTAAGATTTTGAAATTACAGATTCAA 2808
 QY 881 IleMetGlnLysLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnPro 900
 DB 2809 ATATATGAAAGAAATATGAAACCAAGAGAAATATGCTGCAACATTAACAAATTA 2868
 QY 901 SerLeuIleThrPheLeuGlnLysAsnGlnSerValLeuAlaCysSerGlyGlnAspIle 920
 DB 2869 TCTCTAATACCTTCTTCTGCAAAAACTCAGTGTCTGCTGCTGCTGCTGCTGCTGCT 2928
 QY 921 HisValIleGlnLysMetHisHisValAsnMetThrProGluPheLysGlnLeuLysIle 940
 DB 2929 CATGTAATGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2988
 QY 941 ValArgGlnAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlnLysIle 960
 DB 2989 GTACAGAGAAACCAACACCTGCAAAAAAGTGTGCGCATCACTCAATAATTAATTA 3048
 QY 961 IleCysLysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 980
 DB 3049 ATCTGCAAAATGAGCCAGCTTGGGCAACAATGATGATGATGATGATGATGATGATGAT 3108
 QY 981 CysLeuLysIleArgAsnPheValValPheLysAsnAsnSerThrLysLysLysIle 1000
 DB 3109 TGTCTCAAAATATAGCAATTTTGTAGTGTGTTTCAAAAAATTAATTCACAAAGAAAC 3168
 QY 1001 LysLysTrpValGlnLeuProIleThrPheProAsnLeuAspTyrSerGlyCysLys 1020
 DB 3169 AAAAAGTGGTGAATATTAATCAATCAATTTCCCAATCTGATATTCAGAAATGCT 3228
 QY 1021 PheSerAspGlnAsp 1025
 DB 3229 TTTAGTATGAGGAT 3243
 RESULT 2
 AX300832 3372 bp DNA linear FAI 30-NOV-2001
 LOCUS AX300832
 DEFINITION Sequence 1 from Patent WO0185955.
 ACCESSION AX300832
 VERSION AX300832.1 GI:17382110
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 Bahr, G., Cocude, C. and Capron, A.
 TITLE Rh16 polypeptides and its fragments and polynucleotides encoding

Journal: said polypeptides and therapeutic uses
 Patent: WO 0185935-A 1 15-NOV-2001;
 Istac (FR) : INSTITUT PASTEUR DE LILLE (FR)
 Location/Qualifiers
 1. 3372

FEATURES
 source /organism="Homo sapiens"
 /db_xref="taxon:9606"
 155. 3232
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD13049.1"
 /db_xref="GI:17382111"

translation="MSGYSTDENFRYLISCHRAVKNVLOVEPYLDLTLPAYKE
 QIORTVATSGNMAVEELSTLEKGVNHLGWREVEAKRTGSLARNAPELITDL
 PSPFENADEVDLQDLNLOPTLVKLLVRYLVDEKPELLTIEDNRKIAAENNGN
 ESQRELLKRVQENKNSAFNLRLQNNELVOELTNSDSENALENISQYDG
 OVEQDLSTTVQPLEKEVMENSSSESVSLSAEGSVESLDESLGH
 NSNMDSVCTGSDSDENVARASPELDLRLPOMVAPALEKRLITICLPYSG
 KTRVAVYIAKQHLDDKKKASPEKVIIVLYNKVLYVEOLFKERPOFLKKVRYGLSG
 DTOLKISPEPVKSCDITITAOILENSLLENEDMAVOLDSFLLTIDECNTNK
 EAVYNNIMKHITLMOKLNNRLKKNRPVILPQILGLASPEVGATQQAAREHILK
 LCANLDAFTIKTVKENLDQKNQIOEPCKFAIADATNDEPKELILNRIQYCO
 MSPMDFQOPYEOWAIOEKKAEGNKESEVCAENHLKYNKALQINDITIMIDAT
 HLETFYNEEKOKFAVIEDDSDEGDDLYCDDPEDDI KKPLKIDEDREFLMTFFE
 NKKMLKRLAENPEYNEKTLKRLNTIMVYTRFESARLIFTKRYSAVALSOMITE
 NEKRAEYKAKHLIGACHSSEFPKPTONEKEVSKTRGKINLILATVAEELDI
 KRCNTVIRGLVTNEIIVAVQARAHAESTYVIVAHSSQVETIETENDREKMYK
 AIIHVOYNNKPREYAKHILELOMOSIMEKMKRNIJAHYNNPULITPLKNCSTLA
 CSGEDIHVLEKHNHYMTPREKELYIVENKALQKCAIYQINSHIICQCAWCTM
 VHKGLDPLCKLRNFVYVFNKNTKQYKKWELPITFEINDYSRCLSPED"

BASE COUNT 1157 a 637 c 748 g 830 t

Alignment Scores:

Pred. No.: 0 Length: 3372
 Score: 5285.00 Matches: 1020
 Percent Similarity: 99.71% Conservative: 2
 Best Local Similarity: 99.51% Mismatches: 3
 Query Match: 99.51% Indels: 0
 DB: 6 Gaps: 0

US-09-515-363c-2 (1-1025) x AX300832 (1-3372)

QY 1 MetSerAsnGlyTyrSerThrAspGluAsnProPheArgTyrLeuIleSerCysPheArgAla 20
 DB 155 ATGTCCAAATGGGATTTCACACGACGAAATTCCTCCATCTCATCCGCTATTCACAGGCTC 214
 QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuProAla 40
 DB 215 AGGGTGAATAATGTACATCCAGGTGGAGCTGTCTGCTACTTCACTTCTGCTCTGA 274
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetClnAlaValGln 60
 DB 275 GAGGTGAGAGGAGCATTCAGAGGACAGTCGACCTCCGGAACATGACGACAGTGA 334
 QY 61 LeuLeuLeuSerThrLeuGluLysGlyValIlePheIleLeuGlyTyrThrArgIlePheVal 80
 DB 335 CTGCTCTGACGACCTTGGAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 394
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 DB 395 GAGGCCCTCCGGAGAACCCGACCTCTGCGCGCGCGCTACACTGAACTTGAAGCTCAGC 454
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnIleLeuAsnLeu 120
 DB 455 GACTTCCTCTCCATCGTTTGGAGACGCTCATGATGATATGCTCAATTCGCGAACTC 514
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGlu 140
 DB 515 CTTCACCCCACTCTGTGTGACAACTTCTAGTAGACGCTCTTCATATATGTCATGAG 574
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnGlnAsn 160

DB 575 GAGCAACTGTTGACAAATTTGAAGACAAACCGGATTCGTCTGCGAAGAAACATGTA 634
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpPheValAla 180
 DB 635 GAATCAGGTGTAGAGAGACTACTATAAAGATTCGCGAAGAAAGAAACTGGTCTCTGCA 694
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnGlnLeuValGlnGluLeuThrLysSer 200
 DB 695 TTTCGAATGTTCTTCGCAACAGAAACATGATCTGTCACAGATTTACACACTCTCT 754
 QY 201 AspCysSerGlySerAsnAlaGluIleGluAsnLeuSerGlnValAspGluProGluVal 220
 DB 755 GATTCTCAGAAAGCAATGCAAGATTTGAGAAATTTATACAAAGTTGATGTCCTTAACTG 814
 QY 221 GluGlnGlnLeuLeuSerThrThrValGlnProAsnLeuGlnLysGluValIleTrpLysMet 240
 DB 815 GAAGAGCAACTCTTTTCAACCCAGCTTCAGCCAAATCTGGAGAGAGAGCTGGCAATG 874
 QY 241 GluAsnAsnSerSerGluSerPheAlaAspSerSerValValSerGluSerAsnIleThr 260
 DB 875 GGAATTAACATCATCAAGATCATCTTTGCAATCTCTGTAAGTTTCAAGATTCACATCA 934
 QY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlnLysAsnSerAsnMet 280
 DB 935 ACTTTGGCAGAGGAAGAGTCACTGCTTACATGAAAGCTTTGGACATACAGCAATG 994
 QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluAsnValAlaAlaValAla 300
 DB 995 GGCATGATTCAGGACCATGGAGAGTATTCAGATGAAAGAAATGTGCGACAGCAAGCA 1054
 QY 301 SerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProValLeu 320
 DB 1055 TCCTCCGAGCCAGAACTCAGCTCAGGCTTACCAAAAGGAAGTGCCTCCACCACTT 1114
 QY 321 GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVal 340
 DB 1115 GAAGGAGGAAT 1174
 QY 341 TyrIleAlaLysAspHisLeuAspLysLysLysLysAlaSerGluProGlyLysValIle 360
 DB 1175 TACATTGCCAAGACATCACTTACACAAAGAAAGAAAGCATCTGACCTCGGAAAGTATA 1234
 QY 361 ValLeuValAsnLysValLeuLeuValGlnGlnLeuPheArgLysGluPheGlnProThr 380
 DB 1235 GTTCTTGCAATTAAGTACTGCTAGTATACACAGCTCTTCCGAGAGCTTCCACCATTT 1294
 QY 381 LeuLysLysTrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe 400
 DB 1295 TTGAAGAAATGATATGATGATATGATGATGATGATGATGATGATGATGATGATGAT 1354
 QY 401 ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSer 420
 DB 1355 CCAGAAAGTTGCAAGTCTGTGATATTTATCACTACAGCTCAAAATCTTGAAAGACTCC 1414
 QY 421 LeuLeuAsnLeuGluAsnGlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIle 440
 DB 1415 CTCTTAACCTGGAAAGAGGAGAGATCTGCTGTTCAATTCAGACTTTCCTTCATTT 1474
 QY 441 IleIleAspGluCysHisIleIleThrAsnLysGluAlaValIleTyrAsnAsnIleMetArgHis 460
 DB 1475 ATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
 QY 461 TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIlePro 480
 DB 1535 TATTGATGACAGAACTTGAAGAAACATGAGCTCAAGAAAGAAACCAACCGATATCC 1594
 QY 481 LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAla 500
 DB 1595 CTTCCTCAGATACCTGGAGCTAACACCTTCACTGCTGTTGGAGGGGCGACAGCAACCT 1654
 QY 501 LysAlaGluGlnLysIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThr 520
 DB 1655 AAGCTGAAGAACACATTTTAACATATGTCCAATCTTGATGATGATTTACTATTAAACT 1714

QY 521 ValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysPheAla 540
 DB 1715 GTTAAAGAAACCTGATCACTGAAAAACCAATACAGAGCTATGCAAGAGTTGGC 1774
 QY 541 IleAlaAspAlaThrArgGluAspProPheLysGluLysLeuGlnIleMetThrArg 560
 DB 1775 ATTCGAGATGCAACGAGACAGATCCATTAAAGAAACCTTCTGAAATATGACAGC 1834
 QY 561 IleGlnThrCysGlnMetSerProMetSerAspPheLysThrGlnProCysGln 580
 DB 1835 ATTCAAACTATGCAATGAGTCCCAATGCGATTGGAACCTCAACCTATACACA 1894
 QY 581 TrpAlaIleGlnMetLysLysAlaAlaLysLysGluAsnArgLysLysValCys 600
 DB 1895 TGGCCCATTCAAATGAAAAAGCTGCAAAAAGAGAAATGCAAAAGAAAGTGTCT 1954
 QY 601 AlaGluHisLeuArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleArgMetIle 620
 DB 1955 GGAGAACATTGAGAGATGACATAAGCCCTACCAATATACCAATACGAAATGATA 2014
 QY 621 AspAlaTyrThrHisLeuGlnThrPheTyrAsnGluLysLysPheAlaVal 640
 DB 2015 GATGGCTATACCTCATCTTGAACCTTCTTAATGAGACAAAGATTAAGAGTTTGAC 2074
 QY 641 IleGluAspSerAspGluGluGluAspAspGluTyrCysAspGluLysAspGlu 660
 DB 2075 ATAGAGATGATGATGATAGGCTGATGATGATGATGATGATGATGATGATGATG 2134
 QY 661 AspAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhe 680
 DB 2135 GATGATTTAAGAAACCTTGAACAGTGAACAGATGATGATGATGATGATGATGAT 2194
 QY 681 PheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluTyrLysAsnGlu 700
 DB 2195 TTGAAACAAATTAATGTTGAAAAAGCTGCTGAAACCCGAAATATGAAATGAAAG 2254
 QY 701 LeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgThrGluLysLeuAla 720
 DB 2255 CTGACCAAAATTAAGAAATCCCAATAGCAATAGCAATAGCAATAGCAATAGCA 2314
 QY 721 GlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnIleThrGln 740
 DB 2315 GGATATATCTTTACAAAAACGACAGAGATGCATATGCCCTTCCAGAGATTTACTGAA 2374
 QY 741 AsnGluLysPheAlaGluValGlyValLysAlaHisLysLeuIleGlyAlaLysHis 760
 DB 2375 AATGAAAAATTTGCTGAAGTAGAGTCAAAAGCCACCATCTGATGAGTGGACACAG 2434
 QY 761 SerGluPheLysProMetThrGlnAsnGluGlnLysGluValIleSerLysPheArg 780
 DB 2435 AGTGAGTTTAAACCCATGACAGCAAGATGAACAAAGAAATCAATGTAATTTCCACCT 2494
 QY 781 GlyLysIleAsnLeuLeuIleAlaThrThrValAlaGluGluLysAspIleCysGln 800
 DB 2495 GGAATAATTAATCTGCTTATGCTACCAAGTGGCAGAGAGAGTGTGAAATTAAGAA 2554
 QY 801 CysAsnIleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValIleAla 820
 DB 2555 TGTAAATCTGATATCCGATATGCTCTGTCACCAATGAATACCATGAGTGGTGG 2614
 QY 821 GlyArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlySerGly 840
 DB 2615 GGTGCGAGCCAGAGCTGATAGACACCTACGCTCGTGTGCTACAGTGTCTCAGAGTT 2674
 QY 841 IleGlnHisGluThrValAsnAspPheArgGluLysMetMetLysLysAlaIleHisCys 860
 DB 2675 ATGGAAGCGAGACACTATATGATTCGAGAGAAATATGATATGATATGATATGAT 2734
 QY 861 ValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetGlnSer 880
 DB 2735 GTTCAAAATATGAAACCAAGAGATGCTCATGAATTTGGAATTTACAGATCAAGAA 2794

QY 881 IleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAspPro 900
 DB 2795 ATAAATGAAAAAGAAATGAAAAACCAAGAAATATTTGCCAGAGATTTCACAAATTAACGCA 2854
 QY 901 SerLeuIleThrPheLeuCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIle 920
 DB 2855 TCACATAATACCTTCCTTGCAGAAAAACCTGAGTGTGAGCTGTTCGAGGAAATATC 2914
 QY 921 HisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLysPhe 940
 DB 2915 CATGTAATTTGAGAAATGCAATACAGTCAATATGAGCCCGAATTCAGGAATTTATAT 2974
 QY 941 ValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyGluIle 960
 DB 2975 GTAAGAGAAACCAAGCACTGCAAAAGAGTGTGCCAGCTATCAATTAATGAGCAATC 3034
 QY 961 IleCysLysCysGlyGlnAlaTyrPheLysThrMetMetValHisLysGlyLeuAspLeuPro 980
 DB 3035 ATCTGCAATGTGGCCAGCTTGGGAGCAATGATGTGTCACAAAGGCTTACATTTCCCT 3094
 QY 981 CysLeuLysIleArgAsnPheValValAlaPheLysAsnAsnSerThrLysLysGlnTyr 1000
 DB 3095 TGTCTCAAAATTAAGAAATTTGTAAGTGTTCACAAATTAATCAACAAAGAAATATC 3154
 QY 1001 LysLysTrpValGluLeuProIleThrPheProAsnLeuAspTyrSerGluCysLys 1020
 DB 3155 AAAAGTGGTGAATTTACCTATCATCATTTCCCAATCTTGACTATTCAGAAATGCTGTTA 3214
 QY 1021 PheSerAspGluAsp 1025
 DB 3215 TTTAGTATGAGAGAT 3229
 RESULT 3
 AT017378 3373 bp mRNA linear HKI 31-JAN-2001
 LOCUS
 DEFINITION Homo sapiens RNA helicase-DEAD box protein RH116 mRNA, complete cds.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 AUTHORS
 1 (bases 1 to 3373)
 Coccone, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-D., Capron, A. and Bahr, G.M.
 Identification of a new RNA helicase (RH116) regulated by the immunomodulator Murabutide
 TITLE
 JOURNAL
 AUTHORS
 unpublished
 2 (bases 1 to 3373)
 Coccone, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-D., Capron, A. and Bahr, G.M.
 Direct Submission
 Submitted (09-JAN-2001) Laboratoire d'Immunologie Moléculaire de l'Infection et de l'Inflammation, Institut Pasteur de Lille, 1 rue du Professeur Calmette BP 245, Lille 59019, France
 FEATURES
 source
 1..3373
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 155..3232
 /codon_start=1
 /product="RNA helicase-DEAD box protein RH116"
 /protein_id="AA054076.1"
 /db_xref="GI:12621066"
 /translation="MSNGSTDENFRFLISCFRAKVKYIOVEPVLLYTFELPAEYVE
 OIOGTVATSGMVAVELLSTLEGWHLICHTREFVALRTGSPSLAARYMDELTDL
 PPSFENAHDETLDLNLQDTLVLDKLLVRYLDKCEHEELITIEDNRRIAAENKGN
 ESGVRELKRIYQENMFSAFLNVLRTGNNELVQELTGDSSNSAEINLSQVDDP
 QVEQLISTVQPMLEKVEWGMENNSSESPASSVSESDTSLAEQSVLSDLESGLH
 NSNGSPSGTSDSDENVAARASPEPELDRIYQVEVQAPALEGNIIILCPGTGG
 KTRAVYIAKHLDKKKKASPEKAVIYLVNKKYLLVEGLPFKEITPLKKMYRVIIGLSG

QY 581 TTPALAIIEGIMETGLULYSYSLAALATYSLYSGYASARGLYSGLUARVALCY 600
 DB 1895 TGGGCCATTCAATGAAAAAAGCTGCAGAAAAAGAGAAATCGCAAAAGTCTTTCT 1954
 QY 601 AAGIHLHSLAARGLYSTYRANGIUALALEUGINILEASNPPTHLEARGMETILE 620
 DB 1955 GCAGAAACATTGAGGAGTACAAATTAAGGCCCTACAAATTAATGACACATTCGATGATA 2014
 QY 621 ASPALATYTHHISLEUGLUTHRPHEITYRANGIUGLULYSASPYSIYSPHEALV 640
 DB 2015 GATGGCTATACCTCATCTGAACTTCTATAATGAGAAAGATTAAGAACTTTCAGCTC 2074
 QY 641 IIEGLUASPSPERASPGLUGLYASPAPGLUTYCYASNGLYASPGIULASPGLU 660
 DB 2075 ATAGAAAGATGATGATGATGAGGCTGCTGATGATGATGATGATGATGATGATG 2134
 QY 661 ASPASPGLULYSYSPROLEULYSLEUASPGLUTHRASPARGPHELEUMETTHLEUPHE 680
 DB 2135 GATGATTTAAAGAAACCTTGAACTGGATGAACAGATGATTCCTCATCTTATTT 2194
 QY 681 PHEGLUASNPGLULYSMETLEULYSARGLEUALAGIULASNPGLUTYCYASNGLY 700
 DB 2195 TTTGAAAACAAATTAATGTTGAAAAGGCTGCTGAAAACCCGAAATATGAAATGAAAAG 2254
 QY 701 LEUTHRYSLEUARGASNPTRHLEMETGLUGLUTHRARGTHRGTHRGTHRGTHRGTH 720
 DB 2255 CTGACCAATTTAGAAATTCATTAATGAGCATATCTAGAGCTGAGGATGACACAGCA 2314
 QY 721 GLYLIELEPHEHRLYSTRARGINSERATATYALALESEGLINTRPLETHRGTH 740
 DB 2315 GGAAATATCTTTACAAAAACACAGACAGATGATATGCCCTTCCAGAGCTACTGAA 2374
 QY 741 ASNGIULYSPPHAGLULVALGLYVALYSLALAHSHISLEULIEGLYALAGLYHISER 760
 DB 2375 AATGAAAAATTTGCTGAATAGAGTCAAAAGCCACCACTGATTTGGACATGACACAGC 2434
 QY 761 SERGLUPHELYSPROMETHRGINSANGIULYSGLULVALIIESELYSPHEARGTHR 780
 DB 2435 AATGAGTTCAAAACCATGACACAGATGACAAACAAAGATTAATTAATTTCCACT 2494
 QY 781 GLYLYSLIEASNPGLULIEALATHRTHRVALAAGIUGLULYSPLIEULYSGLU 800
 DB 2495 GAAAAAATTAATCTGCTTATCTGCTACCAAGTGGCAAGAGGCTGATTTAAGAA 2554
 QY 801 CYASNPILLEVALIIEARGYRGLYLEUVALIHRASNGIULIEALAMEVYIENALARG 820
 DB 2555 TGTACATTGTTATCCGTTATGCTCTGTCACCAATGAATAGCCATGTCAGGCCCT 2614
 QY 821 GLYATGALIAARGALASPGLUSERTHRYVALLEUVALIAHISSEGLYSERGLYVAL 840
 DB 2615 GGTCCAGACGAGCTGATGAGACACTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2674
 QY 841 IIEGLUHLISGLUTHRVALASNPHEARGGLULYSMETMETTYRYSALAHISCY 860
 DB 2675 ATCGAAACGAGACAGTTAATGATTCGAGAGAGATGATTAATTAATTAATTAATTA 2734
 QY 861 VALGINASMETLYSPROGLUGLUTYRALAHISLYSILEUGLULEUGLIMETGLINSTR 880
 DB 2735 GTTCAAAATATGAAACCAAGAGATGCTCATAGATTTCGAAATTACATGCAAAAT 2794
 QY 881 IIEMETGLULYSMETLYSTRYASARGASNPILIEALYSHISTYLSASNPSPRO 900
 DB 2795 ATATATGAAAAAAGAAATGAAAAACCAAGAAATATTTCCACACATTAACATTAATTA 2854
 QY 901 SERLEULIETHRPHLEUCYSLYASNCYSERVALLEUALACYSERGLYGLULASP 920
 DB 2855 TACTATTAATTAATCTCTCTGCAAAAACTGACAGTGTCTGCTGCTGCTGCTGCTGCTG 2914
 QY 921 HISVALIIEGLULYSMETLISHVALASNPMTHTHPRGGLUPHELYSGLULEUTHYLE 940
 DB 2915 CATGTAATGAGAAAAATGATCATGCTCATGATGATGACCCCAAGATTCAGAGAACTTACAT 2974

QY 941 VALARGIULASNPALALEUGINLYSLYSCYSLAASPTRYGINILEASNPGLULIE 960
 DB 2975 GTAGAGAAAAAAGCACTCTCAAAAGAGTGTGCCACTTCAATTAATTAATTAATTAAT 3034
 QY 961 IIECYLSYSCYGLYGINALATRPGLYTHRMETVALHISLYSGLYLEUASPSPRO 980
 DB 3035 ATCTCAATGTGGCCAGCTTGGGGAACATGATGTGTCACAAAGGCTTACATTTGCT 3094
 QY 981 CYSEULYSILEARGASNPHEVALYVALPHELYSSASNPSPRTHRYSTYSPHITYR 1000
 DB 3095 TGTCTCAAAATTAAGAAATTTGTAGTGTCTTCAAAATAATTAATTAACAAAGAAATTA 3154
 QY 1001 LYSYSTRPVAGLULIEUPROILETHRPHROASNPULASPYRSEGLYCYSEYLEU 1020
 DB 3155 AAAAAGTGGTAGAATTAATCTCATCATTTCCCAATCTTGACTTACGAATAGCTGTTA 3214
 QY 1021 PHESEASPGLUASP 1025
 DB 3215 TTTAGTATGAGAGAT 3229
 RESULT 4
 AY075132
 LOCUS 3078 bp mRNA linear 4 20-MAY-2002
 DEFINITION Mus musculus HELICARD mRNA, complete cds.
 ACCESSION AY075132
 VERSION AY075132.1 GI:18698980
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 3078)
 AUTHORS
 Kovacsovics, M., Mattinson, F., Michéau, O., Bodmer, J., Hofmann, K., and Tschopp, J.
 TITLE
 Overexpression of Helicard, a CARD-Containing Helicase Cleaved during Apoptosis, Accelerates DNA Degradation
 JOURNAL
 CURR. BIOL. 12 (10), 838-843 (2002)
 MEDLINE
 22009979
 PUBMED
 12015121
 REFERENCE
 2 (bases 1 to 3078)
 Kovacsovics, M., Hofmann, K., and Tschopp, J.
 DIRECT SUBMISSION
 Submitted (24-JAN-2002) Biochemistry, University of Lausanne, ch
 des Boveresses 155, Epalinges, VD 1066, Switzerland
 FEATURES
 source
 1..3078
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1..3078
 /note="Helicase; contains two CARD domains"
 /codon_start=1
 /product="HELICARD"
 /protein_id="FAL77205.1"
 /db_xref="GI:18698981"
 /translation="MSIVCSAEDSRMLILFPRRLKMYTOVEVLHLLFLSAETKE
 OIKKINTCGNTSAELLSTLEGOMPLMTOMFVAFHLSGTVLAARVAKPTLDL
 PPSRELRHDECHLLTLLOPTLDKLLINDVLDTCERGLITVEPRNRLISAAGNCP
 ESGVRELRRLIVOKENPFSTFLDVRKLDNDALFOELTGCGCPQNDNLANSRQCP
 AANECCLIPAVDESSLETEAVNNVDDIIDEASCTDSSVTESTSI AEGSVCPFESSLGH
 NSNMRGDSGTMSDSDSVIOTKRVSPPELQIRPOMEVAQPAIDKNI11CLPTGS
 GKTAVAYITKDLHDKKQASGSKVIVLVNKAVALAOLPFKFNLYLKMYRLIGS
 GDLOKLSPEPVYKSYDVIISTAGIILNSLINSLEGGDDVQI SPSSL11IDSCHTN
 KEAYNNIMRRLYLOKRLNNDIKONKPAIPILOTIGLTAISPVGAARKQSEKHLI
 NTCANLDAFTITKVENUGOLKHOIKRCPKPFYADDTRENPKKILETMASTOTYC
 OKSPMSDFGTQHEQMAIOMEKKAQGNKRDVCAHEKRYEALQINDTIRMIDAY
 SHLETPTDERKEKFAVLNDSKSDDEASSCNDQKGVKSKILETDFLNLDFD
 NKKMLKILAKPKYENKELIKRLTILEQPTRESESSNGI11FKTOSTYALQWIME
 NKAFAEYVAHHLIGAGHSSEVPMTOTOKEYISKFRGEINL11ATVAEGEIDI
 KECNIVIRGLYNETIAMOARGAARDESTYVLTSSGVTREKIVNDRERKNYK
 AINRVOAKKPEEVAHKLLELOVOSTILEKKKRVKSTAKQYNDN11ITLCKKSMIY
 CSGENHIVIEKMHVNMTPPEKGLYVIREKKAOLKFRADYVUTN11ICKGQAMGTM
 VHKGLDLPCKIRNFVYVFNNSPKKQYKMWELPIRFPOLDYSHVCLYSDSD"
 BASE COUNT 1012 a 644 c 715 g 707 t

ORIGIN

Alignment Scores:

Pred. No.: 1,58e-277 Length: 3078
 Score: 4199.00 Matches: 818
 Percent Similarity: 87.04% Conservative: 75
 Best Local Similarity: 79.73% Mismatches: 131
 Query Match: 79.06% Indels: 2
 DB: 10 Gaps: 2

US-09-515-363c-2 (1-1025) x AY075132 (1-3078)

QY 1 MetSerAsnGlyThrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20
 DB 1 ATCTCGATTGCTGTTCTCGACGACAGCTTCAGAACTCATCTTATTTTCACGCCCT 60
 QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrIleThrPheLeuProAla 40
 DB 61 AGGCTGAATAATGATACATTCAGGTGGAGCCAGCTGCTGACCACCTCATCTTCTGTCTGTA 120
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGln 60
 DB 121 GAACCAAAAGAGCAGATTCTTAAAGATCAACACCTGTGTAAACACCAATCGGAGCA 180
 QY 61 LeuLeuLeuSerThrLeuGluLysGluValAlaThrPheIleGlnTyrThrArgIlePheVal 80
 DB 181 CTGCTCTGACGACCTTGAGGAGCAATGCCCTGTGGATGGAGCAGCATGTTCTGCGTG 240
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 DB 241 GAGCGCTGAGACAGAGTGGCAATCCCTGAGCGCGGCTATGTCAAATCTACACTCAT 300
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuAsnLeu 120
 DB 301 GATCTCCCTCTCTCTCTCTCTGAGACTGACAGCATGAGTGTCTCCACTTCTGACCCCTG 360
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGlu 140
 DB 361 CTCACCCCACTTGTGTGACAACTTGTGATTAAACAGATGCTTGGACATGCTTCTGCG 420
 QY 141 GluGlnLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnGlnIleAsn 160
 DB 421 AAGGACATTTGACATCGACAGCAAGAAATCCGATTTCCTGCTCAGGAATACGGGAAAT 480
 QY 161 GluSerGlyValArgIleLeuLeuLysArgIleValGlnLysGluAsnThrPheSerAla 180
 DB 481 GAGTCAGGTGTAAGAGAGTGTGAGACAGATTGTGCACAGAGAAACATGTTTCTATC 540
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGlnLeuValGlnIleThrGlySer 200
 DB 541 TTCCTGTGATGTTCTCGCCAACTGGAATGATGCACTTCCAAAGAACCTAACAGCTGCA 600
 QY 201 AspCysSerGluSerAsnGluIleGluAsnLeuSerGlnValAlaSerProGlnVal 220
 DB 601 GCGTGGCCAGAAACACACACAGACTTGCTAACTGCTGCACAGAGATGAGCTGAGAG 660
 QY 221 GluGlnGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValAlaProGlyMet 240
 DB 661 AATGATGCTGTTCTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 241 GluAsnAsnSerSerGluSerPheAlaAspSerSerValValSerGlnSerAspThr 260
 DB 721 GAGGACATATTACCAAGAGCTTCTGTACAGATTCGTGTGATGACCAACAAATCACACACA 780
 QY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlnHisAsnSerAsnMet 280
 DB 781 AGTTTGGCAGAAAGAGTGTACCTCTCTGATGAAAGTCTTGAGCATATACAGCAACATG 840
 QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaAlaArg 299
 DB 841 GCGAGGATTCAGGACCATGGAAGTGTATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 300 AlaSerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAla 319

DB 901 GTATCCCGGACGACGAACTGAGCTCAGGCTTACCAAAATGAAATGGCTCAATTCAT 960
 QY 320 LeuGlnGlyLysAsnIleIleGlyLeuProThrGlySerGlyLysThrValAla 339
 DB 961 CTAGATGGGAAGAAATATATATATATATCTGCTCCCGACGGGAGTGGGAAACGAAATGCT 1020
 QY 340 ValTyrIleAlaLysAspHisLeuAspLysLysLysAlaSerGluProGlyVal 359
 DB 1021 GTTTCATCATCCCAAAAGATCATTACACAAAGAGGAGGATGCTGAATCCGAGAAATTT 1080
 QY 360 IleValLeuValAsnLysValLeuValGlnGlnLeuPheArgLysGluPheGlnPro 379
 DB 1081 ATCGTTCTGTCATTAAGCATTAATGTTACAGACAACTTTCCGAAAGAGTCAATCCCA 1140
 QY 380 PheLeuLysLysTyrTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSer 399
 DB 1141 TATTGAAAGAAATGATATGATATTTGATGATTAATGTCGATATCCACCTGAAATATCA 1200
 QY 400 PheProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsn 419
 DB 1201 TTTCCAGAAATGTTCAAAATCTTACATGTTATTCACACCTGCTCAAAATCCTTAAAC 1260
 QY 420 SerLeuLeuAsnLeuGluAsnGlyLysAspAlaGlyValGlnIleSerAspPheSerLeu 439
 DB 1261 TCCCTCTTAAATCTGAGAGTGGAGACATGACGCTGTGACGCTGTGACGCTGTGCTC 1320
 QY 440 IleIleIleAspGluCysHisIleThrAsnLysGluAlaValTyrAsnAsnIleMetArg 459
 DB 1321 ATTATCATTTGATGATGTCATGACACACCAAGAGGAGCTATTAACATCATCATGAGA 1380
 QY 460 HisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIle 479
 DB 1381 CGATATTGTAAGCAGCAACACTGAGAAACATGACCTCAAGAAACAAACCAACCAAT 1440
 QY 480 ProLeuProGlnIleLeuGlnLeuThrAlaSerProGlyValGlnGlyAlaThrGlySer 499
 DB 1441 CCCCTGCGGAGATCTAGAGACTGACACTTACCTGCTGTGAGACAGCCAAACAGAG 1500
 QY 500 AlaLysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleCys 519
 DB 1501 TGTGAGCTGTAAGAAACATATTTAAATATATGTCGCAATCTTGAATTCATTAATA 1560
 QY 520 ThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysPhe 539
 DB 1561 ACAGTGAAGAGAAATCTGTCACTCAACTCAACCAACCAAAAGAAACCTGCAAGAAATTT 1620
 QY 540 AlaIleAlaAspAlaThrArgGluAspProPheLysGlnLysLeuGlnIleMetThr 559
 DB 1621 GTGATTGCTGATGACACACGAGAAATCAATTTAAAGAGAAATCTTGAAGAAATTAAGCA 1680
 QY 560 ArgIleGlnThrTyrCysGlnMetSerProMetSerAspPheGlyThrGlnProTyrGln 579
 DB 1681 AGCATTCAGACTTACTGCAACCAAGGAGGATGATGATGATGATGATGATGATGATGATGAT 1740
 QY 580 GlnTyrAlaIleGlnMetGluLysLysAlaAlaLysLysGlnLysAsnArgLysGluProVal 599
 DB 1741 CAGTGGCCATTCATTAATGAGAAAGAAAGCTGCTAAAGCGGAGAAATCCGAAAGATGAGCT 1800
 QY 600 CysAlaGlnHisLeuArgLysTyrAsnGlnAlaLeuGlnIleAsnAspThrIleArgMet 619
 DB 1801 TGTGAGAAACATTTGAGAGATGACAGACGACCTTCAAAATCAAGACAGCATCCCATG 1860
 QY 620 IleAspAlaTyrThrHisLeuGlnIleThrPheTyrAsnGlnLysAspLysLysPheAla 639
 DB 1861 ATGATGATATATGACACTGAGACATTTACACTGATGAGAAAGAAAGAAATGATGCA 1920
 QY 640 ValIleGluAspSerSerAspGluGlyLysAspGluTyrCysAspGlyAspGluAsp 659
 DB 1921 GTCTC---AATGACAGGACAGAGAGTATGACAGGACGACACTTGCATTAATGCAATTT 1977
 QY 660 GluAspAspLeuLysLysProLeuLysLeuAspGlnThrAspArgPheLeuMetThrLeu 679

| | | | |
|----|------|--|-------|
| Db | 1978 | AAGGGCGATGTAAGAAATCTTTGAAACCTGGAGCAAGCAACGATGAATTCCTATGAAATTTC | 203 |
| QY | 680 | PhepGciuaSnAsnLysMetLeuLysATGLeuAlaGluAsnProGluTY:GluAsnGlu | 699 |
| Db | 2038 | TTCTTTTCAACCAAGAAATGTTGAAAAAACTACAGTGAAGAACCCAAAAATACGAGATGAA | 20997 |
| QY | 700 | LysLeuThrLysLeuArgAsnThrIleMetGluGlnTYrThrArgThrGluGlnSerLysA | 719 |
| Db | 2098 | AAATCTATTAAATTAAGAAACACGATACGGAACAATTCACAAAGCTGGACGAGCTCCGC | 2157 |
| QY | 720 | ArgGlyIleIlePheThrLysThrArgGlnSerAlaTYrAlaLysSerGlnTriPleThr | 739 |
| Db | 2158 | CGAGAGATTATTTTTCACAAAAACAGACAGAGAGACTACGACCTTCCACATGATCATG | 2217 |
| QY | 740 | GluAsnGluLysPheAlaGluValGlyValLysAlaHisHisLeuIleGlyValaGlyHis | 759 |
| Db | 2218 | GAATATCCAAAGTTTGCGGAAGTTGGAGTCAAAAGCCACACCTGATTCCCGCGGAGAC | 2277 |
| QY | 760 | SerSerGluThrLysProMetThrGlnAsnGluGlnLysGluValIleSerLysPheArg | 779 |
| Db | 2278 | AGCAGTAAGTCAAAGCCCATGCTGACACTGAAACAAAAAGAGTATTGATAAATTTCGC | 2337 |
| QY | 780 | ThrGlyLysIleAsnLeuLeuIleAlaThrThrValAlaGluGluGlyLeuAsnTriLys | 799 |
| Db | 2338 | ACTGCGCAATTAATTCGCTATTCGCTACGACGGTGGCAACAGAAAGCCATGATATCCAA | 2397 |
| QY | 800 | GluCysAsnIleValIleArgTYrGlyLeuValThrAsnGluIleAlaMetValGlnAla | 819 |
| Db | 2398 | GAGTGCATATTGTTATTCGTTATGCGCTTGTCACAAACAGATAGACCATGCTCAGATC | 2457 |
| QY | 820 | ArgGlyArgAlaArgAlaAspGluSerThrTYrValLeuValAlaHisSerGlySerGly | 839 |
| Db | 2458 | CGGGCTGAGCCAGACGCTGATGAAGACACGACTATGCTCTGTACACAGCACTGGCTCACAA | 2517 |
| QY | 840 | ValIleGluHisGlnThrValIleAsnAspPheArgGluLysMetMetTYrLysAlaIleHis | 859 |
| Db | 2518 | GTTACCGAAGCGGAGATGTTGTTATGATTCGCCAGAAACAAAGATGATGATTAAGCTATTAC | 2577 |
| QY | 860 | CysValGlnAsnMetLysProGluGluTYrAlaHisLysTleuGluGluGlnMetGln | 879 |
| Db | 2578 | CGTGTTCACAAACATGAACACAGAGAGATATGACATTAAGATTGGAATTCACGTCGCAA | 2637 |
| QY | 880 | SerIleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisTYrLysAsnSn | 899 |
| Db | 2638 | AGTATCCCGAAAGAAAGAAATGAAGTCMAAAGAGCATTCMAAGCAATATACGACACAT | 2697 |
| QY | 900 | ProSerLeuIleThrPheLeuCysLysAsnCysSerValIleuAlaCysSnGlyGluLys | 919 |
| Db | 2698 | CCATGCTTAATACACTTCTCTGCAAAAATTTGATGATG:TTGCTTGCTTGCGAAGAAAC | 2757 |
| QY | 920 | IleHisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLeuTYr | 939 |
| Db | 2758 | ATTCATGTCATTTGAAAGATGATGCATCATGTCATATATGCAACAGAAATTCAGCTTAC | 2817 |
| QY | 940 | IleValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTYrGlnIleAsnGlyGln | 959 |
| Db | 2818 | ATTGTAAAGAGAAACAAAGACACTGCCAAACAAATTTGCTATATATCAGATGCAATGAGAG | 2877 |
| QY | 960 | IleIleCysLysCysGlyGlnAlaTYrGlyThrMetMetValHisLysGlyLeuAsnPhe | 979 |
| Db | 2878 | ATTATCTCCAAAGTGTGGCGAGGCTGGGGAAACATGATG:TTGCACAAAGCTTTAGATTGG | 2937 |
| QY | 980 | ProCysLeuLysIleArgAsnPheValValValPheLysAsnAsnSerThrLysLysGln | 999 |
| Db | 2938 | CCTGTCTTAAATTAAGAGAAATTTTGTTGTCATATTTCAAAAATTAACGCACATCAAGAAACAG | 2997 |
| QY | 1000 | TYrLysLysTYrPValGluLeuProIleThrPheProAsnLeuAspTYrSerGluCysCys | 1019 |
| Db | 2998 | TACAAAGAGTGGTGGAATGCTCATTCGATTTCCCTGAGTTTGACTACTACTGCAATACTCC | 3057 |
| QY | 1020 | LeuPheSerAspGluLys 1025 | |
| Db | 3058 | TTGTATGATGATGAAGAT 3075 | |

```

Db 235 ATGTGCGATTGCTGTTGCGAGGAGGAGCTTACAGAAATTCATTCTTCAGGCTC 294
OY 21 ArgValLysMetThrIleGlnValGluProValLeuAspIleThrPheLeuProAla 40
Db 295 AGGCTGAATATGATACATTCAGTGGAGGAGGAGCTGCTGAGCACTCATCTTTCGTCTGA 394
OY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu 60
Db 355 GAAACCCAAAGAGAGATTTCTTAAAGATCAACCTGTGTATCAACCAACGAGGAGGAGAA 414
OY 61 LeuLeuLeuSerThrLeuGlnLysGlyValThrPheLeuLeuThrPheArgGluPheVal 80
Db 415 CTGCTGCTGAGCACTTGGAGGAGGAGCAATGGCTCTGTGATGAGGAGGAGATGTTGCTG 474
OY 81 GluAlaLeuArgThrGlySerProLeuAlaAlaArgThrMetAsnProGluLeuThr 100
Db 475 GAGGCGCTAGAGCAGAGTGGCAATCCCTTAGCGCGCGGCTATGTCAAAGTCACACTGACT 514
OY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluThrLeuGlnLeuLeuAsnLeu 120
Db 535 GATCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 594
OY 121 LeuGluProThrLeuValAspLysLeuLeuValArgAspPheLeuAspLysGlyMetGlu 140
Db 595 CTCCAGCGCCACTTGGGAGCAAACTTGATTAACGATCTCTTGACATTCCTCTCGAG 654
OY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaLeuLysAsnGlnIleAsn 160
Db 655 AAGGACTATTGACAGTGCAGAGCAGAGCAAAATCGGATTTCTGCTGCGAGAAATACGGGAT 714
OY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGlnAsnThrPheSerAla 180
Db 715 GATCTAGGCTTAAAGAGAGCTGCTGAGAGAGATTTGTCAGAGGAGAAACTTCTTTCTAC 774
OY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrGlySer 200
Db 775 TTCTGTGATTTCTGCGCCAACTGGAATGATGCACTATTCACAAATCAACAGCTGCA 814
OY 201 AspGlySerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAlaAspIleProGlnVal 220
Db 835 GCGTGGCCAGCAAGCAACAGACAGCTTGGCACTGCTGCTGACAGAGATGAGTGTGAGCT 894
OY 221 GluGluGlnLeuLeuSerThrValGlnProAsnLeuGlnLysGluValTrpGlyMet 240
Db 895 AATGACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
OY 241 GluAsnAsnSerSerGluSerPheAlaAspSerSerValValSerGluSerAspThr 260
Db 955 GAGCAGATATTACAGAGGCTTTCTTACAGATTCGCTGTGACTACAGATTCAGAGCA 1014
OY 261 SerLeuAlaGluLysSerValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMet 280
Db 1015 AGTTGGCAAGAGAGTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
OY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluValAlaAlaValArg 299
Db 1075 GGGAGGAGTTCAGGCACTGAGAGTGTGATGCTGAGTGAAGTGTCTCCACCAAAAGAA 1134
OY 300 AlaSerProGluProGluLeuGlnLeuArgProThrGlnMetGlnValGlnGlnProAla 319
Db 1135 GTATCCCGCCAGCAGAACTGCACTGCTGAGGCTTACCAATGATGATGCTCAACAGCT 1194
OY 320 LeuGluGlyLysAsnIleIleCysLeuProThrGlySerGlyLysThrArgValAla 339
Db 1195 CTGATGGGAGAGATTTATTTATCTGCTCCCGCCAGGAGGAGTGGCAAAACGAGAGCT 1254
OY 340 ValThrIleAlaLysAspHisLeuAspLysLysLysAlaSerGluProGlyLysVal 359
Db 1255 GTTACATCATCCAAATCATCTTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1314
OY 360 IleValLeuValAsnLysValLeuLeuValGluGlnLeuPheArgLysGlnPheGlnPro 379
Db 1315 ATGCTTCTGTCATTAAGTAAATGTTAGCAGAACAACTTTCCGAAAGAGTTTAACCA 1374
OY 380 PheLeuLysLysThrPheArgValIleGlyLeuSerGlyAspThrGlnLeuThrLeuSer 399
Db 1375 TATTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434
OY 400 PheProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsn 419
Db 1435 TTTCGAGAGTGTCAAACTTACGATGTTATTATCAGACAGCTGCAAACTCTTAAGAC 1494
OY 420 SerLeuLeuAsnLeuGlnAsnGlyGlnAspAlaGlyValGlnLeuSerAspPheLeuSer 439
Db 1495 TCCCTTTAAATCTGAGAGTGAAGAGATGAGAGTGTGAGAGTGTGAGAGTGTGAGAGT 1554
OY 440 IleIleLeuAspGluCysHisHisThrAsnLysGluAlaValThrAsnAsnIleMetArg 459
Db 1555 ATTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1614
OY 460 HisThrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIle 479
Db 1615 CGATATTGACAGAGAGCTGAGAAACATGACCTCAAGAAACAAACAAACAAACCAACAT 1674
OY 480 ProLeuProGlnIleLeuGlnLysLeuThrAlaSerProGlyValGlyGlyAlaThrLysGln 499
Db 1675 CCGCTGCGCAGATCTAGAGCTGACAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1734
OY 500 AlaLysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLys 519
Db 1735 TGTGAGCGCGAAACATATTTTAAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1794
OY 520 ThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnLysProCysLysLysThr 539
Db 1795 ACAGTGAAGAGAAATCTGTGCTCAACCTCAACCAACCAACCAACCAACCAACCAAT 1854
OY 540 AlaIleAlaAspAlaThrArgGluAspProPheLysGlyLysLeuGlnLeuLeuThr 559
Db 1855 GTGATTTGCTGATGACACCAAGAAATCTTAAAGAAATCTTAAAGAAATCTTAAAGAA 1914
OY 560 ArgIleGlnThrTyrcGlnMetSerProMetSerAspPheGlyThrGlnThrProGlyLys 579
Db 1915 ACCATTGACAGCTTACTGCGCAAAAGTCCAAATGCTCAATTTTGGAAACCAATATATGAG 1974
OY 580 GlnThrAlaIleGlnMetGluLysLysAlaAlaLysLysGlyAsnArgLysGluArgVal 599
Db 1975 CAGTGGCGCATCAATATGAGAAAAAAGCTGCTAAAGAGCGAAATCGCAAAATATATGCT 2034
OY 600 CysAlaGluHisLeuArgLysThrAsnGluAlaLeuGlnIleAsnAspThrIleArgMet 619
Db 2035 TGTGCAAGACATTTGAGAGAGTACACAGAGCGCTTCAAAATCAACGACAGCTTCAATG 2094
OY 620 IleAspAlaArgThrHisLeuGlnThrPheTyraGlnGluLysAspLysLysHisAla 639
Db 2095 ATTGATGATATAGCAGCTGAGGCTTCTTACACTGATGAGAAAGAAAGAAAGAAAGTCC 2154
OY 640 ValIleGluAspAspSerAspGluGlyLysAspAspGluThrCysAspGlyAspThrLys 659
Db 2155 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2211
OY 660 GluAspAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeu 679
Db 2212 AAGGCGGATGTAACAAATCTTTGAACCTGAGCAAGAGAGAGATTTCTCATGATTTG 2271
OY 680 PhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluThrGluAsnGlu 699
Db 2272 TTCTTTGATTAACAAAGAAATGTTGAAAAAAGTACGTGAAAAACCAAAATACGAGATGAA 2331
OY 700 LysLeuThrLysLeuArgAsnThrIleMetGlnGlnThrArgThrGlnGluSerAla 719
Db 2332 AAACATCAATTAATTAAGAAACAGTACTGCAACATTCACAAAGTGTGAGAGAGCTCC 2391
OY 720 ArgGlyIleIlePheThrLysThrArgLysSerAlaArgLysGlnThrIleThr 739
Db 2392 CGAGGAATTAATTTTACAAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 2451

```

[illegible]

QY 515 Alaphethlilelysthrvalylsgluasnleuasplnleuylsasnleuilectnlu 534
 |||||
 Db 121 GGATTACTATTAAACCTTTAAAGAAACCTTGATCAACCAAACTAAATACAGAG 180
 |||||
 QY 535 ProCysLysPheAlaIleAlaAspAlaThrArgGluAspProPheGlnLysLeu 554
 |||||
 Db 181 CCATGCAGAAATGTTGCCCTTCAGATGCACACAGAGAGATCCATTAAAGAGAACTT 240
 |||||
 QY 555 LeuGluIleMetThrArgIleGlnThrTyrcysGlnMetSerProMetSerAspPheGly 574
 |||||
 Db 241 CTAGAAATATATGCAAGAGATCAAACTTATGTGCAAAATGATGTCATGTTTGGCA 300
 |||||
 QY 575 ThrGlnProTyrcysGlnThrAlaIleGlnMetGluLysAlaAlaLysGln 594
 |||||
 Db 301 ACCCAACCTATGACAAATGGCCATTCAAATGAAAAAGAGAGAGAGAGAGAGAG 360
 |||||
 QY 595 ArgLysGluArgValCysAlaGluHisLeuArgLysTyrAsnGluAlaIleGlnLeu 614
 |||||
 Db 361 CGCAAGAAACGCTGTTGTGCAGACATTTGAGAGAGAGAGAGAGAGAGAGAGAG 420
 |||||
 QY 615 AspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheTyrAsnGluLys 634
 |||||
 Db 421 GACACATTCGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 |||||
 QY 635 AspyLysPheAlaValIleGluAspAspSerAspGluGlyGlyAspAspLys 654
 |||||
 Db 481 GATTAAGAACTTGTGCACTATAGAAAGATGATGATGATGATGATGATGATGAT 540
 |||||
 QY 655 AspGlyAspGluAspGluAspLysLeuLysProLeuLysLeuAspGluThrAsp 674
 |||||
 Db 541 GATGCGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 |||||
 QY 675 PheLeuMetThrLeuPhePheGlnLysAsnLysMetLeuLysArgLeuAlaGluAsnPro 694
 |||||
 Db 601 TTTCCTCATGACTTATTTTAAAAACAATAAATGTTGAAAGAGCTGGTGAAGAAACCGTA 660
 |||||
 QY 695 GluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArg 714
 |||||
 Db 661 GATATGAAATGAAATGAAAGTGAACCAATTAAGAAATACCATATGAGTAATATCTAGG 720
 |||||
 QY 715 ThrGluGluSerAlaArgGlyIleIlePheThrLysThrArgGluSerAlaTyrAlaLeu 734
 |||||
 Db 721 ACTGAGAGATGACAGCAGAGAGATTAATCTTACAAAACGACAGAGAGAGAGAGAG 780
 |||||
 QY 735 SerGlnTyrIleThrGluAsnGluLysPheAlaGluValGlyValLysAlaHisIleu 754
 |||||
 Db 781 TCCCACTGATTAAGTAAATGAAATTTGCTGAAGTACGAGTCAAAAGTCAACCATCTG 840
 |||||
 QY 755 IleGlyAlaGlyHisSerSerGluPheLysProMetThrGlnAsnGluGlnGluVal 774
 |||||
 Db 841 ATTGGAGCTGACACAGCAGTGAAGTCAAAACCATGACACAGAAATGAATAAAGAGAGTC 900
 |||||
 QY 775 IleSerLysPheArgThrGlyLysIleAsnLeuLeuAlaThrThrValAlaGluGlu 794
 |||||
 Db 901 ATTAGTAATTTTGGCAGCTGGAAGAAATTAATCGCTTATCGTACACAGCTGGCAGAGAAA 960
 |||||
 QY 795 GlyLeuAspIleLysGluCysAsnIleValIleArgTyrGlyLeuValIleAsnGluLe 814
 |||||
 Db 961 GGTCTGGATATTAAAGAAATGTAACATTTGATCCGTTATGCTGCTCACTCAATGAATA 1020
 |||||
 QY 815 AlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluSerThrTyrValIleuValAla 834
 |||||
 Db 1021 GCCATGGTCCAGCCCGTGGTGGAGCCAGAGCTGATGAGAGAGAGAGAGAGAGAGAG 1080
 |||||
 QY 835 HisSerGlySerGlyValIleGluHisGluThrValAsnAspPheArgGluLysMetLys 854
 |||||
 Db 1081 CACAGTGGTTCAGAGATTAATGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGATGATG 1140
 |||||
 QY 855 TyrLysAlaIleHisCysValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeu 874
 |||||
 Db 1141 TATAAAGATACATGCTGCTCAAAATATGAAGAACAGAGAGATGCTCAATGAAGATTTG 1200
 |||||
 QY 875 GluLeuGlnMetGlnSerIleMetGluLysLysMetLysThrLysArgAsnIleAlaLys 894

Db 1201 GAATTACAGATGCAAAAGATATATGAGAAAGAAATGAAACCAAGACAAATATATATGAC 1260
 |||||
 QY 895 HisTyrLysAsnAsnProSerLeuIleThrPheLeuCysLysAsnCysSerValAla 914
 |||||
 Db 1261 CATTCAGAAATTAACCCATCACTATAACTTCCCTTCGCAAAAACGTCAGAGCTGTACGC 1320
 |||||
 QY 915 CysSerGlyGluAspIleHisValIleGluLysMetHisHisValAsnMetThrProGlu 934
 |||||
 Db 1321 TGTCTGGGAGAAATATCATGTAATGAGAAATGATGATGATGATGATGATGATGAT 1380
 |||||
 QY 935 PheLysGluLeuTyrIleValArgGluAsnLysAlaLeuGlnLysCysAlaAspLys 954
 |||||
 Db 1381 TTCAGAGAACTTACATTTGTAAGAGAAACAAACACGCAAGAGAGAGAGAGAGAGAG 1440
 |||||
 QY 955 GlnIleAsnGlyGluIleIleCysLysCysGlyGlnAlaIleArgPylThrMetMetValHis 974
 |||||
 Db 1441 CAATTAATATGATGAATATATGCAAAATGTCGCCAGGCTTGGGAGACATATGATGTCAC 1500
 |||||
 QY 975 LysGlyLeuAspLeuProCysLeuLysIleArgAsnPheValValPheLysAsnAsn 994
 |||||
 Db 1501 AAAGCTTAGATTTGCTGCTGCTCAAAATTAAGAAATTTTGTAGTGGTGTTCACAAATTAAT 1560
 |||||
 QY 995 SerThrLysGlnTyrLysTyrIleValGluLeuProIleThrPheProAsnGluAsp 1014
 |||||
 Db 1561 TCACAAAGAAACCAATACAAAGAGGAGTGAATTAATCTATACATATTCACAAATTTGAC 1620
 |||||
 QY 1015 TyrSerGlyCysCysLeuPheSerAspGluAsp 1025
 |||||
 Db 1621 TATTCAGAAATGCTGTTTATTTAGTATGATGAGAT 1653
 |||||

RESULT 7
 LOCUS BC025508 2468 bp mRNA linear 5' end AUG-2002
 DEFINITION Mus musculus, clone MGC:38064 IMAGE:5252257, mRNA, complete cds.
 ACCESSION BC025508
 VERSION BC025508.1 GI:19343862
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogamathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2468)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lohrer Hennighausen Ph.D., Robin Humphreys CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (C.N.)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland.
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Bakesley, R.W., Boufard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Leaspi, R., Maduro, Q.L., Masfello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://imgc.llnl.gov>
 Series: IRAK plate: 56 Row: c Column: 12
 This clone was selected for full length sequencing because it

Db 1382 AAGATAGAGAGTTCAGTACATAGAGATGATGAGCTGGTTCATGATGATAT 1441

RESULT 9
BC004031 2050 bp mRNA linear ROD 07-AUG-2002

LOCUS
DEFINITION Mus musculus, clone IMAGE:3495361, mRNA, partial cds.

ACCESSION BC004031

VERSION BC004031.1 GI:13278455

KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 2050)
Strausberg, R.
Direct Submission
Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

AUTHORS
TITLE
JOURNAL

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINI)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mdickpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES
source
Location/Qualifiers
1..2050
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZCH 11"
/clone="IMAGE:3495361"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
<1..1545
/codon_start=1
/product="Unknown (protein for IMAGE:3495361)"
/protein_id="AA04031.1"
/db_xref="GI:13278455"
/translation="ANDAFITIKYENIGQIKHQIKKCKFVIADDPREPKEL
LEIMASITTCQSPMSDFGTHTEWMAIQMERKAKNDNRDRAAEHUKYNELO
INDTIRMDAYSHELEFYDEKREKPAVINDSDSDPSNDLQKGVKSKLKE
TDEFLNLFENKMKLKLAEKPKYENIKLIRNTILQFRESSESGIIFITRQ
STYALSOIMENAKFAEVGVKAKHLIGAGHSEVKPMTLQEKVEYSKRTEINIL
ATTVAEEGLDIKEONIVIRYGLVETLAWOARGRAESTYVLYTSAGSVTEERI
VNDREKMYKAIINRYQNNKPREYAKLELVQSLPDKKMYKSTIAQYNDNEEL
TLCKKCSMLVSGENIHVIEKMHAVNIPERGLIVTINKALOKKRDYDTNGEII
CRGQAMGTMMVHKGLDLPCLIRNFVNFKNNSPKRYKKKVELPIRPDDISYC
LYSDSD"

CDS
BASE COUNT 747 a 374 c 427 g 502 t

ORIGIN

Alignment Scores:
Pred. No.: 3, 87e-146 Length: 2050
Score: 2271.50 Matches: 438
Percent Similarity: 90.87% Conservative: 30

Best Local Similarity: 85.05% Mismatches: 46
Query Match: 42.77% Indels: 1
DB: 10 Gaps: 1

US-09-515-363c-2 (1-1025) x BC004031 (1-2050)

OY 511 AlaAsnLeuAspAlaPheThrIleLeuThrValIysGluAsnLeuAspGlnLeuYasn 530
|||||
DB 1 GCCAATCTTGATGCTCTTACCATTTAAACAGTGAAGAAATCTTGGTCACTTAAACAC 60
|||||

OY 531 GlnIleGlnIleProCysLeuLysPheAlaIleAlaAspAlaThrArgGluAspProIle 550
|||||
DB 61 CAATTAAGCAACCATGCAAGAAATTTGATGCTGATGATACCAAGAAATCAAT 120
|||||

OY 551 LysGluLysLeuLeuGluIleMetThrArgIleGlnThrYrCysGlnMetSerProMet 570
|||||
DB 121 AAAGAGAAACCTAGTAAATTTGGCAGCATTCAGACTTCAGCCAAAAGGTCATG 180
|||||

OY 571 SerAspPheGlyThrGlnProTyrGlnIleProIleAlaIleGlnMetGluLysIleAla 590
|||||
DB 181 TCAGATTTGGAAACCCACCATTTATGACGACGTGGCCATTCAAATGACAAATACTGCT 240
|||||

OY 591 LysLysGlyAsnArgLysGluArgValCysAlaGluHisLeuArgLysTyrAsnProIle 610
|||||
DB 241 AAAGAGCAAAATCGCAAAAGATGCGCTGTGGCAAAATTTGAGCAAGATACAAACCC 300
|||||

OY 611 LeuGlnIleAsnAspThrIleArgMetIleAspAlaThrHisLeuGluThrIleThr 630
|||||
DB 301 CTACAAATCACGACACGATCCGAATGATGATGGCTATAGCCACTGACACACATCTAC 360
|||||

OY 631 AsnGluGluLysAspLysLysPheAlaValIleGluAspAspSerAspGluThrLys 650
|||||
DB 361 ACTGATGACAAAGAAAGAAAGATTCGACGCTCC---AACGACAGGACGACGACGAC 417
|||||

OY 651 AspGluTyrCysAspGlyAspGluAspGluAspLeuLysLysProLeuLysLeuAsp 670
|||||
DB 418 GAGGCGCAGAGTGGCAATGATCCCACTTAAGGGCGATGTAAACAAATCTTTGAAATG 477
|||||

OY 671 GlnThrAspArgPheLeuMetThrLeuPheGluAsnAsnLeuMetLeuLysArgLeu 690
|||||
DB 478 GAACGAGATGAATTTCTCTGTAATTTGTTCTTGATACACAAATGTTGAATAACATA 537
|||||

OY 691 AlaGluAsnProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGlu 710
|||||
DB 538 GCTGAAAACCCCAAAATACGACGATGAAACAACTTAATTAAGAACGACGATATGGA 597
|||||

OY 711 GlnTyrThrArgThrGluGluSerAlaArgGlyIleIlePheThrLysThrArgIleSer 730
|||||
DB 598 CAATTCACAAAGCTGACGAGCTCCCGAGGAATATTTCACAAAACGACGACGACGAC 657
|||||

OY 731 AlaTyrAlaLeuSerGlnThrIleThrGluAsnGluLysPheAlaGluValGluLys 750
|||||
DB 658 ACCTAGCAGCTTCCAGGATCATGAAATGCAAAATTTGGCGAATTTGAATCCAA 717
|||||

OY 751 AlaHisIleLeuIleGlyAlaGlnHisSerSerGluPheLysProMetThrGlnAsn 770
|||||
DB 718 GCCGATCAGCTGATTTGCCCGGCGACACGACGATGAAGTAAACCCATGCTAGACT 777
|||||

OY 771 GlnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeuLeuIleAlaThr 790
|||||
DB 778 CAATAAGCAATCTATTGTAATTTCCGACTGCGCAATTAATCTGTTATCTGATGACG 837
|||||

OY 791 ValAlaGluGluGluLysAspIleLysGluLysAsnIleValIleArgGlyGlyLeu 810
|||||
DB 838 GTGGCAGAGAAAGCGCTGATATCAAAAGATGCAATATTGTTATTTGTTATGCTT 897
|||||

OY 811 ThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaAspGluSerThrIleThr 830
|||||
DB 898 ACGAATGAGATGACCATGCTGCTCCAGCGCGGGGTGAGCCAGACGATGATAAAGCAG 957
|||||

OY 831 ValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAspPheArg 850
|||||
DB 958 GTCCGTGATACCGACGAGTGGCTCAGAGATTACCGAAGGAGGATGTTAATGATTCG 1017
|||||

QY 292 AspgIUGluAsnValAlaAlaIarGAlaSerProGluProGluLeuGlnuArGProTyr 311
 DB 671 GATGAAGAAATGTGGACCAAGACATCCCGGAGCCGAAAGTCCAGCTTAC 730
 QY 312 GlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleCysLeuProThr 331
 DB 731 CAAATGGAAGTGGCCAGCCAGCTTGGAAAGGAATATCATCATCTCCCTACA 790
 QY 332 GlySerGlyLysThrArgAlaAlaValTyrIleAlaLysAspHisLeuAsnLysLys 351
 DB 791 GGAAGGAAAAACCAAGAGTGGCTGTTTACATTGCCAAGATCAGCTTACAAAGAAAGAA 850
 QY 352 LysAlaSerGluProGlyLysValIleValLeuValAsnLysValLeuLeuValGluGln 371
 DB 851 AAAGCATCTGACCTGGAAAGTAACTAGTCTTGTCAATAAGCTACTGCTATGTAACAG 910
 QY 372 LeuPheArgLysGluPheGlnProPheLeuLysLysTyrTyrArgValIleGlyLeuSer 391
 DB 911 CTCTCCGCAAGAGAGTCCACCATTTTGAAGAAATGGTATCGTTATGGATTAACT 970
 QY 392 GlyAspThrGlnLeuLysIleSerPheProGluValIleLysSerCysAspIleIleIle 411
 DB 971 GGTGATACCACTGAAATATCATTTCCAGAAAGTTGTCAAGTCCGTGAATTATTATC 1030
 QY 412 SerThrAlaGlnIleLeuGluAsnSerLeuLeuAsnLeuGluAsnGlyGluAspAlaGly 431
 DB 1031 AGTACAGCTCAATCTCTTAAAGTCCCTTAAGCTTGAATAATGAAAGATGCTGCT 1090
 QY 432 ValGlnLeuSerAspPheSerLeuIleIleIleAspCysHisIleThrAsnLysGlu 451
 DB 1091 GTTCAATCTCAGACTTCCCTCATTTATCATGATGATGATGATGATGATGATGATGAT 1150
 QY 452 AlaValTyrAsnAspIleMetArgHisTyrLeuMetGlnLysLeuLysAsnArgLeu 471
 DB 1151 GCAAGTATATATACATCATGAGCATTTATTGATGCAAGATGAAAAAAGAAAAA 1210
 QY 472 Lys 472
 DB 1211 AAA 1213
 RESULT 11
 AX300838 1284 bp DNA linear PAT 30-NOV-2001
 LOCUS AX300838
 DEFINITION Sequence 7 from Patent WO0185955.
 ACCESSION AX300838
 VERSION AX300838.1 GI:17382116
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1
 Bahr, G., Cocude, C. and Capron, A.
 Rh16 polypeptides and its fragments and polynucleotides encoding
 said polypeptides and therapeutic uses
 Patent: WO 0185955-A 7 15-NOV-2001;
 JOURNAL Istac (FR) : INSTITUT PASTEUR DE LILLE (FR)
 FEATURES
 source Location/Qualifiers
 1..1284
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 489 a 198 c 261 g 335 t 1 oth: s
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,68e-127 Length: 1284
 Score: 1995.00 Matches: 378
 Percent Similarity: 99.47% Conservative: 0
 Best Local Similarity: 99.47% Mismatches: 2
 Query Match: 37.56% Indels: 0
 DB: 6 Gaps: 0
 US-09-515-363c-2 (1-1025) x AX300838 (1-1284)

QY 646 AspgIUGluGlyAspAspGluTyrCysAspGlyAspGluAspGluAspLeuLysLys 665
 DB 2 CATGAGGGTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 61
 QY 666 ProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhePheGluAsnAsnLys 685
 DB 62 CTTTGAAGTGGATGAACAGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
 QY 686 MetLeuLysArgLeuAlaGluAsnProGluTyrGluAsnGluLysLeuThrLeuArg 705
 DB 122 ATGTTGAAGAGCTGGCTGAAACCCAGAAATATGAAATGAAAGCTGACCAATATAGA 181
 QY 706 AsnThrIleMetGlnGlnTyrThrArgThrGlnGlnGlnSerAlaArgGlyIleIlePheThr 725
 DB 182 AATACCATTAATGAGCAATATATCTAGGCTAGGCAATAGCAGCAGCAATATATCTTACA 241
 QY 726 LysThrArgLysSerAlaTyrAlaLeuSerGlnTyrIleThrGluAsnGluLysPheAla 745
 DB 242 AAACACGACAGAGTGCATATGCGCTTCCAGTGCATTAAGTAAATGAAAAATTGCT 301
 QY 746 GluValGlyValLysAlaHisHisIleLeuIleGlyAlaGlyHisSerSerGluPheLysPro 765
 DB 302 GAAGTAGAGAGTCAAGCCACCATCTGATTTGAGCTGACACAGAGAGTTCATAAACC 361
 QY 766 MetThrGlnAsnGluGlnGluValIleSerLysPheArgThrGlyLysIleAsnLeu 785
 DB 362 ATGACACAGATGAACCAAAAGAAAGTCAATTAATTTCCAGCTGCAAAAAATATATG 421
 QY 786 LeuIleAlaThrThrValAlaGluGluGluLysAspIleLysGluCysAsnIleValIle 805
 DB 422 CTTATCGCTACACAGTGGCAAGCAAGCTCGATTTAAAGATTAACATTTGATGATGATGAT 481
 QY 806 ArgTyrGlyLeuValIleThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAla 825
 DB 482 CTTATGCTCTGCTCACCAGTAATGACCAATGAGCTGACGAGGCGGCTGACGACGAGACT 541
 QY 826 AspGluSerThrTyrValLeuValAlaHisSerGlySerGlyValIleGlnHisGlnThr 845
 DB 542 GATGAGAGCAGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
 QY 846 ValAsnAspPheArgGluLysMetMetTyrLysAlaIleHisCysValGlnAsnMetLys 865
 DB 602 GTTAATGATTTCCGAGAGCAATGATGATTAAGCTATATACATTTGTTCAAAATATGAAA 661
 QY 866 ProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetGlnSerIleMetGluLysLys 885
 DB 662 CCAGAGAGTATGCTCATTAAGATTTTGAATTAAGATGCAAGATTAATGCAAAAGAAA 721
 QY 886 MetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnProSerLeuIleThrPhe 905
 DB 722 ATGAAACCAAGAGAAATTTTCCCAAGCATTTCAAGAAATACCCATCACTAATATCTTC 781
 QY 906 LeuCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIleHisValIleGlnLys 925
 DB 782 CTTTGAAGAACTGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
 QY 926 MetHisHisValAsnMetThrProGluPheLysGluLeuTyrIleValArgGluLysLys 945
 DB 842 ATGCATCAGCTCAATATGACCCCAAGAAATTAAGAACTTACATTTGTAAGAGAAAGAAA 901
 QY 946 AlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyGluIleIleCysLysCysLys 965
 DB 902 ACACCTGCAAAAGAGTGTCCGACATCAATTAATTTGGTGAATATATCTGCAAAATGT 961
 QY 966 GlnAlaTyrGlyThrMetMetValHisLysGlyLeuAspLeuProCysLeuLysIleArg 985
 DB 962 CAGGCTGTGGAAACATATGATGTGCACAAAGCTTATGATTTCCCTGCTCAAAATATAGG 1021
 QY 986 AsnPheValValAlaPheLysAsnAsnSerThrLysLysGlnTyrLysLysTyrValIle 1005
 DB 1022 AATTTGTATGTTGTTTTCAAAAATATATTCACAAAGAAACCAATATCAAAAGAGGCTTA 1081

QY 1006 Leuprolidehphroanleuaptyrserglucyscysleuphscysaspgluasp 1025
 DB 1082 TTACCTATCATCTTCCCATCTTGCATTCAGATGCTGTTATTATTATGATGAGAT 1141
 RESULT 12
 AF316999
 LOCUS AF316999 2326 bp mRNA linear R00 04-MAR-2001
 DEFINITION Mus musculus D1119p2 mRNA, complete cds.
 VERSION AF316999.1 GI:13194585
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2326)
 Miyoshi, K., Cui, Y., Riedinger, G., Robinson, P., Lebecky, J.,
 Zou, L., Oka, T., Dewar, K. and Hennighausen, L.,
 Structure of the mouse Stat 3/5 locus: evolution from Drosophila to
 zebrafish to mouse
 TITLE
 JOURNAL Genomics 71 (2), 150-155 (2001)
 MEDLINE 21100887
 PUBMED 1161808
 REFERENCE 2 (bases 1 to 2326)
 Cui, Y., Zhang, M., Hanover, J. A. and Hennighausen, L.,
 Identification of two genes in the Stat3/5 locus that encode novel
 proteins located in the endoplasmic reticulum and the cytoplasm
 UNPUBLISHED
 JOURNAL 3 (bases 1 to 2326)
 Cui, Y. and Hennighausen, L.,
 Direct Submission
 TITLE Submitted (24-OCT-2000) Laboratory of Genetics and Physiology,
 National Institutes of Health, 8 Center Drive, Bethesda, MD 20892,
 USA
 FEATURES
 source
 1. 2326
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /map="next to Gcn5"
 1. 2326
 /gene="D1119p2"
 2. 2038
 /gene="D1119p2"
 /codon_start=1
 /product="unknown"
 /protein_id="AAK15474.1"
 /db_xref="GI:13194585"
 translation="MELRPYQMEVILPALKENKILIMLIGAGKTRAAAFYAKRHLET
 VDRGVVYLVNRYVHLVSHAEFRMLDKHPTVTLLSCKSRGFGIMARSHDLIC
 TAEILQIALNSSEDEHELEFSLIVDECHHKHKTIVNTLSRYLEOKLRKEPL
 PVDLQLTASPGTGATKLGALIDHILQICANIDTCINSIPKNCYSOLIMHPKCKQY
 DLQORRADPFGDLIKKILNQIHQOLEPDLKQCGTQMYEQVAVOLCKDAEAGLQE
 ORVYALHLRYNDALFIDHTVARADALMDLQVDRERTKTOVWRAESMLLKLPDC
 KNVGOLARGPENKLEMLERILLKQCSCHRGILFETRTQASLSLLMLKQCP
 LQTVGIRKPMILGAGNTSOSTHMTOKDOEYI JERDRIISLVAISVAEGCLDAQC
 NVVATGILINDISWQAKGRARAGQSVSEIATEGSEKMRRLVNLNLEVLAKRAVA
 AVQKMDPEFRFAKIRIDQOASLVKRAAAVARRHILQGGHILPEHVQLDLCINMVAVG
 SDLRVVEGTHVNNVNPFSVYTTISUNVAVINKEFKDKKKGITIRGSCSEVMQFMI
 YKSVTLPLVAKIGSMLETPRCKIIOAKKSRVSPVPIIIODCTOSLSLSLID"
 BASE COUNT 591 a 594 c 655 g 485 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,44e-83 Length: 2326
 Score: 1356.00 Matches: 299
 Percent Similarity: 59.69% Conservative: 129
 Best Local Similarity: 41.70% Mismatches: 237
 Query Match: 25.53% Indels: 52
 DB: 10 Gaps: 2
 US-09-515-363c-2 (1-1025) x AF316999 (1-2326)

QY 306 LeuGlnLeuAlaArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyAsnLeu 325
 DB 2 ATGGAGAGCGCGACCCCTACAGAGTGGAGTACTTACCTGCTGGAGGCAACAAATATC 61
 QY 326 IleIleCysLeuProThyGlySerGlyLysThrArgValAlaValTyrTlAlaLysAsp 345
 DB 62 ATTATATGGCTGCCCGACCGCTGCTGGCAACACCGCGAGCTGCTTGTATATCAACAGG 121
 QY 346 HisLeuAspLysLysLysAlaSerGluProGlyLysValIleValLeuValAsnLys 365
 DB 122 CATCTAGAG-----ACGGTACACAGAGCGAAGGTGGTGGTACTGTCATACAGG 169
 QY 366 ValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProPheLeuLysLysIlePyr 385
 DB 170 GTACACCTGTGAGGACAG---CACGCCAGGAGTTCAGGCCCATGCTGATATACAG 226
 QY 386 ArgValIleGlyLeuSerGlyAspPheGlnLeuLysIleSerPheProGluValValLys 405
 DB 227 ACTGTAAACACCTGAGTGGGACACATGCCAGTGCCTTGGCTGATGCTGG 286
 QY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLeuAsnGln 425
 DB 287 ACCCAGACCTGCTCATCTGTACGACGACAGTGTACAGTGGCAGCTAACATATCTAG 346
 QY 426 AsnGlyLysAlaGlyValGlnLeuSerAspPheSerLeuIleIleAspGluCys 445
 DB 347 -----GAGATGAGAACACGTAGAGCTCAGAGATTCGTGATGTGTGACAGTGT 400
 QY 446 HisHisThrAsnLysGluAlaValTyrAsnAsnIleMetAlaGlyIleLeuMetGlnLys 465
 DB 401 CACCACACCCACAGAGACACCGCTCTACACACCATCTTGACCCGCTACTGAAAGAGAG 460
 QY 466 LeuLysAsnAsnArgLeuLysLysGluAsnLysProValIleProLeuProIleLeu 485
 DB 461 CTGAG-----AAGCAGAG-----CCCTGCCCTGCTG 493
 QY 486 GlyLeuThrAlaSerProGlyValGlyValAlaThrLysGlnAlaLysAlaGlnLys 505
 DB 494 GGTCTCACAGGCTCCCGGAGCAGTGGAGGCGCACCAAGGCGCATATATCAG 553
 QY 506 IleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThrValLysGluAsnLeu 525
 DB 554 ATCTCAAGCTTTGTGCCAATTTGGATACGTGCCATCATCTGCGCCAAACATATCTAG 613
 QY 526 AspGlnLeuLysAsnGlnIleGlnLeuProCysLysLysPheAlaIleAlaAspAlaThr 545
 DB 614 TCCACAGCTGTGATGCATACCAACCAAGCCCTGCAGGAGTATGACCTGTGAAAGCGC 673
 QY 546 ArgGluAspProPheLysGluLysLeuLeuGlnIleMetThrArgIleGlnIleLysCys 565
 DB 674 GCACAGGATCTTTGGGGAGCTTGATAAAAGCTTATGAACCAATCCAACTCACTA 733
 QY 566 GlnMetSerPromet-----SerAspPheGlyThrGlnProTyrGlnIleThrAlaIleGln 584
 DB 734 GAGATGCTGACTTACACCAACATTTGAAACCCAGATGTATGACCAACCAATGATACAG 793
 QY 585 MetGluLysAlaAlaLysLysGlyAsnArgLysGluArgValCysAlaGlnLysLeu 604
 DB 794 TTGTCCAGAGATGCGCGAGAGGCTGACTCCAGAACAGCGGTGATAGCCTGATTG 853
 QY 605 ArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaLysThr 624
 DB 854 CGGCGCTACATGATGCGCTATTATCCACGATGCTGCGCGCGGCTTGTAC 913
 QY 625 HisLeuGlnIlePheTyrAsnGluGluLysAspLysLysPheAlaValIleGluAsp 644
 DB 914 ATGTTCAGAGATTTTACGACAGAGACGACCAACAAACACAGATGTGCTGCTGAA 973
 QY 645 SerAspGluGlyLysAspAspLysLysCysAspLysAspGluAspAspLysLys 664
 DB 974 AGC----- 976
 QY 665 LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhePheCysAsn 684

```

Db 977 -----TGGCTGCTGCAACGCTTTTTCGACCATCAA 1006
Oy 685 LysMetLeuLysArgLeuAlaGluAsnProGluTyrGluAsnGluLysLysThrLysLeu 704
Db 1007 AATGTGCTGGGCGAGTACAGTACCTGGGCTCTCT---GACAAACCAAGTTCAGATGCTG 1063
Oy 705 ArgAsnThrIleMetGluGlnTyrThrArgThrGluGluSerAlaArgThrIleIleIleIle 724
Db 1064 GAAAGATGCTTACTAGACAGATT---GGGAGTCCCTGGCCACACAGTGGGATCATCTCTC 1120
Oy 725 ThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTyrIleThrGluAsnGluLysPhe 744
Db 1121 ACCAGAACCCGTCAGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1180
Oy 745 AlaGluValGlyValLysAlaHisLysIleGlyAlaGlyIleSerSerGlnPheLys 764
Db 1181 CAGACTGTGGGATCAAGCCGCGAGATGCTGATGGAGGAGAACACAAACCAAGCAGCA 1240
Oy 765 PrometThrGlnAsnGluGlnLysGluValIleSerLysPheArgThrGlyLysIleAsn 784
Db 1241 CACATACCCAGAAAGACAGACAGAGATGATCCAGAGATTCAGAGATGATTCCTCCAGAC 1300
Oy 785 LeuLeuIleAlaThrThrValAlaGluGluLysLeuAspIleLysGluLysAsnIleVal 804
Db 1301 CTCTGTGTCGCAACAAAGTGGCAGAGAGGAGGAGGAGATGATGATGATGATGATGATG 1360
Oy 805 IleArgTyrGlyLeuValThrValThrValIleAlaMetValIleAlaArgThrArgAlaArg 824
Db 1361 GTGGCTATAGGCTCTGTCACCAATGATGATGATGATGATGATGATGATGATGATGATG 1420
Oy 825 AlaAspGluSerThrTyrValIleValIleValIleHisSerGlyValIleLeuHisGln 844
Db 1421 GCGGTCACAGATGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1480
Oy 845 ThrValAsnAspPheArgGluLysMetMetTyrLysAlaIleHisCysValGlnAsnMet 864
Db 1481 CTAAACCAATGAGCTCTGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 1540
Oy 865 LysProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetClnSerIleMetGluLys 884
Db 1541 GACCGTGATGATGATCAAGGCCAGATCGGAGCTTCCACCAACCAATCTCTCTCTCTCTCT 1600
Oy 885 LysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnProLysLeuIleThr 904
Db 1601 GCAGCAGCGCGCGCCATGGGAGATCCAGACAGGCGAGTCTTACCGGATGACGTCGCA 1660
Oy 905 PheLeuCysLysAsnCysSerValLeuAlaCysSerGlyLysAspIleHisValIleGln 924
Db 1661 CTCTCTGATCAACGTATGCGTGGCGGCTGAGGAGTACGGAGTACCTGGCAAAATGTGAG 1720
Oy 925 LysMetHisHisValAsnMetThrProGluPheLysGluLeuTyrIleValArgGluAsn 944
Db 1721 GGCACCCACACGATGATGAAACCCCAACTTCTCGGTACTACTACTACTACTCTCCGAGAC 1780
Oy 945 Lys---AlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnProLysGluIleLys--- 962
Db 1781 CCGTGTGATTTAACAAGGCTTTAAGGACTGGAGACTGGAGCAACCAACAGTCCACT 1840
Oy 963 LysCysGlyValIleAlaTyrGlyThrMetMetValHisLysGlyLysAspIleProCysLeu 982
Db 1841 AACTGTGGGAGGCTGTGGGCTTCCAGATGATGATCAAAATCACTAACCTTGGCAGTGTG 1900
Oy 983 LysIleArgAsnPheValValValPheLysAsnAsnSerThrLysLysGlnTyrLysLys 1002
Db 1901 AAAATGGGAAGC-----ATGCTACTGGAACACCTCGAGGAGAGATCTAGGCCAAAAA 1954
Oy 1003 TrpValGluLeuProIleThrPheProAsnLeuAspTyrSerIleLysCys 1015
Db 1955 TGGTCCCGGGGCTTCTTCATACCAAGTCTTCTGATATCTCTGCA3AAGCTTAC 2005

```

RESULT 13
BC029209

```

LOCUS      BC029209      2336 bp      mRNA      linear      Ref: 07-AUG-2002
DEFINITION Mus musculus, DNA segment, Chr 11, Lohar Hennighausen 2,
ACCESSION   BC029209
VERSION     BC029209.1
KEYWORDS    MGC.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2336)
AUTHORS     Strausberg, R.
TITLE       Direct Submission
JOURNAL     Submitted (01-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
Email: cga@bbs.femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIN)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu/cdna/
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Mizny, D.M.,
Richards, S., Gibbs, R.A.
FEATURES             source
     source          1..2336
     /organism="Mus musculus"
     /db_xref="taxon:10090"
     /map="CECH 11"
     /clone="MGC:35613 IMAGE:2651254"
     /tissue_type="Mammary tumor metastasized to lung.
MMTV-LTR/Int1 model. Expression driven by an MMTV-LTR
enhancer."
     /clone_11b="NC1-CGAP-Lu30"
     /lab_host="DH10B"
     /note="Vector: pCMV-SPORT6"
     13..2049
     /codon_start=1
     /product="DNA segment, Chr 11, Lohar Hennighausen 2,
expressed"
     /protein_id="AAH29209.1"
     /db_xref="GI:20809730"
     /db_xref="LOCUSID:80861"
     /translation="MELRPYMEVILPALEKNILIMLPAGACKAAAPYAKHLET
VDKGVVAVLVKRVHLSCHAEFRMLDKHTVITLSDMSRSGSLMARSHDLITC
TALLQIALNSSEDEHVELEFSLIVDECHHTHTKIVNTILSKYLEOKLKAEP
POLVGLTASPTGATKLOGALIDILQICAMLDNCHLMSFNKYSGLIMHPKPKQXY
DLCORADDPFGDLIKKLMNOIHQOLEPDLKQOFGOMYQOQVVCCKDAEAGLOE
ORVALHILRRYRDLFTIHVYRARDALMDQFDYDREITKTQVWAKESMLKLFDOH
KNVIGOLIAKRPDNPKEMLERLILDKQFSGHRTGRTITRTPTASSLLMLRQOC
LQVGLIKPKMLIGAGNTSOSTHMTQKQEVTOEPRDGLISLVATVAEGDLIAOC
NAVYKADPEFAKIFDLOQALVLRRAAARHREIQOQGFLEPHVOLLICINQVAVGQ
SDLRVECTHHVNPNSVYTTSONPVVINKYFKDMRPQGTIRCSNCGEVMQPMI
YKSVTLPLKIGSMLETPRPCKIQAKKSRVPFIPFDLIQDCTGSLSELSD"
BASE COUNT      590 a      600 c      660 g      486 t
ORIGIN
1
Alignment Scores: 1.44e-83      Length: 2336
Pred. No.:

```

Score: 1356.00 Matches: 299
Percent Similarity: 59.69% Conservative: 129
Best Local Similarity: 41.70% Mismatches: 217
Query Match: 25.53% Indels: 52
DB: 10 Gaps: 12

US-09-515-363c-2 (1-1025) x BC029209 (1-2336)

OY 306 LeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGlnTyrLysAsnIle 325
DB 13 ATGACCTGCGACCGCTACAGGAGGAGATCTTACGCTGCTGAGGCGAAGAATATC 72

OY 326 IleIleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrIleAlaLysAsp 345
DB 73 ATTATATGCTGCGCCACGGGCTGCTGGAGACCGCGCGCTTCTTACCCAAAGAG 132

OY 346 HisLeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 365
DB 133 CATCTAGAG-----ACGGTAGACAGAGCGACAGCTGCTGCTACCTCATTAAG 180

OY 366 ValLeuLeuValGlnGlnLeuPheArgLysGlnPheGlnProPheLeuLysTyrPyr 385
DB 181 GTACACCTGCTGAGCCAG---CACGGCCAGAGATTCAGGATGCTGATTAACACTGG 237

OY 386 ArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPheProGlnValLys 405
DB 238 ACCGTACAAACCTGAGTGGGAGCATGGAGTCCGCTGCTGCTGCTGCTGCTGCTG 297

OY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLysLeuGln 425
DB 298 AGCCACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 357

OY 426 AsnGlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLysAspGln 445
DB 358 -----GAGGATGACACCTAGAGCTCAGAGAAATTCCTGATGTGTGCGACGAGTCT 411

OY 446 HisHisThrAspLysGlnValAlaValTyrAsnAsnIleMetArgHisLysLeuMetGln 465
DB 412 CACACACCCACCAAGGACACCGCTACCAACACCTTACCGCGTACCAACAGAG 471

OY 466 LeuLysAsnAsnArgLeuLysGlnAsnLysProValIleProLeuProGlnIleLeu 485
DB 472 CTGAAG-----AAGGCAAG-----CCCTCTCTGAGGCTCG 504

OY 486 GlyLeuThrAlaSerProGlnIleGlyAlaThrLysGlnAlaLysLysGlnIle 505
DB 505 GGCTCCACGCTCCCGACGACCTGGAGGGGCGACCAAGCTCCAAAGGGGATGATGATG 564

OY 506 IleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThrValLysGlnLeu 525
DB 565 ATCTACACCTTGTGCGAATTTAGATACGTCACACATGATGCTCCCAAGATTCCTGAC 624

OY 526 AspGlnLeuLysAsnGlnIleGlnLysProCysLysPheAlaIleLeuAspAlaThr 545
DB 625 TCCACACTGCTGATGATTAACCGGAAGCCCTGCAAGCAGATGACCTCTCTCAAGCTGC 684

OY 546 ArgGluAspProPheLysGlnLysLeuLeuGlnIleMetThrArgIleGlnThrTyrCys 565
DB 685 GCACACGATCTTTGGGAGCTTGAATAAAAGCTTATACCAACCAATTCACCAACACTA 744

OY 566 GlnMetSerProMet-----SerAspPheGlyThrGlnProTyrGlnIleAlaIleGln 584
DB 745 GAGATGCTGCTGACTGAGCAACAAATTGGACACCAACATATGACAGCAAGCTGATGAC 804

OY 585 MetGluLysLysAlaAlaLysLysGlnAsnArgLysGlnArgValCysAlaGlnHisLeu 604
DB 805 TTGTGCAAGAGATCGGCGACAGGCTGAGCTCCAGAACAGTGGGTGTATGCTGCTGATTTG 864

OY 605 ArgLysTyrAsnGlnAlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaTyrThr 624
DB 865 CGCGGCTACATGATGCTGCTATTTATCCAGTACTGCTGCGCGGAGTCCCTTAC 924

OY 625 HisLeuGlnThrPheTyrAsnGlnLysLysAspLysLysPheAlaValIleGlnLysAsp 644

DB 925 ATGTTGCAAGATTTTACAGACAGACACGACCAACCAACCAACAGATGCTGCTGCTG 984

OY 645 SerAspGlnGlyGlyAspAspGlnTyrCysAspGlyAspGluAspGluAspLeuLys 664
DB 985 AOC----- 987

OY 665 LysProLeuLysLeuAspGlnThrAspArgPheLeuMetThrLeuPheGluAsnAsn 684
DB 988 -----TGGCTGTGAAGGTGTTGATGACATAA 1017

OY 685 LysMetLeuLysArgLeuAlaGlnAsnProGlnTyrGlnAsnGlnLysLeu 704
DB 1018 AATGCTGCGGCGACCTGACAGCTGGGCTGCT---GAGAACCCGAACTTGGAGATGCTG 1074

OY 705 ArgAsnThrIleMetGlnGlnTyrThrArgThrGlnGlnLysLeuAlaGlyIleIlePhe 724
DB 1075 GAAAGATCTTACTAAGCAGTTT---GGAGTCTGCGCCACACTGGGCTATATCTTC 1131

OY 725 ThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTyrIleThrGlnAsnGlnLysPhe 744
DB 1132 ACCACACCCGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191

OY 745 AlaGluValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGluPheLys 764
DB 1192 CAGACTGCGGCGATCACACACAGATGCTGATCGAGCGAGGAGAACACAAAGCAACAGCA 1251

OY 765 ProMetThrGlnAsnGlnLysGlnValIleSerLysPheArgThrGlyLysIleAsn 784
DB 1252 CACATGACCCCAAAAGACACAGAGAGATGATCCAGAGCTTCAGAGATGATGCTGATGAC 1311

OY 785 LeuLeuIleAlaThrThrValAlaGlnGlnLysLeuAspIleLysGlnCysAsnIleVal 804
DB 1312 CTTCAGTGGCGCACAAAGTGTGGAGAGAGGAGGCTGATGCTGCTGCTGCTGCTGCTG 1371

OY 805 IleArgTyrGlyLeuValThrAsnGlnIleAlaMetValGlnAlaArgGlyArgAlaArg 824
DB 1372 GTGCGTATGCGCTCTACCAATGATGATCTCCAGTGGCGCGCGGCTGCTGCTGCTG 1431

OY 825 AlaAspGlnSerThrThrValLeuValAlaHisSerGlySerGlyValIleGlnHisGln 844
DB 1432 GCTGTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1491

OY 845 ThrValAsnAspPheArgGlnLysMetMetTyrLysAlaIleHisCysValGlnAsnMet 864
DB 1492 CTAAACCAATGAGGCTGCTGAGTGTGCTGATGAGAAAGCTGCTGCTGCTGCTGCTG 1551

OY 865 LysProGlnGlnTyrAlaHisLysIleLeuGlnLeuGlnMetGlnSerIleMetGlnLys 884
DB 1552 GACCTGTAGATGATTCAGGCGCAAGATCCGAGCTGCGCAAGCATCTCTGTTAAGCGG 1611

OY 885 LysMetLysThrLysArgAsnIleAlaLysHisLysTyrLysAsnAsnProSerLeuIleThr 904
DB 1612 GCAGACGCGGCGCGCCATCGGAGATCCAGACAGGCGAGCTTCCTACCGAGACCTGTA 1671

OY 905 PheLeuCysLysAsnCysSerValLeuAlaCysSerGlyGlnAspIleHisValIleGln 924
DB 1672 CTCTGCTCATCAATGATGATGCTGCGCTGAGGCTGAGGAGTACCTGCGCAAGATGAG 1731

OY 925 LysMetHisHisValAsnMetThrProGlnPheLysGlnLeuTyrIleValArgGlnAsn 944
DB 1732 GGCACCCACACGCTCAATGATACCCCAACTTCTGCTACTATACCACTCCCAAC 1791

OY 945 Lys-----AlaLeuGlnLysCysAlaLysPyrGlnIleAsnGlyGlyIleCys--- 962
DB 1792 CCTGTGCTATTAACAGGCTCTTAAGGACTGAGACCTGAGAGAACCATAGTGTAGT 1851

OY 963 LysCysGlnGlnAlaTyrGlyThrMetMetValHisLysGlyLeuAspLeuProCysLeu 982
DB 1852 AACTGTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1911

OY 983 LysIleArgAsnPheValValIlePheLysAsnAsnSerThrLysLysGlnTyrLysLys 1002
DB 1911 ----- 1911


```
Db 2063 AACCTGGGGAGGCTCTGGGGCTGCAGATGATCTACAAAGTGAAGCTTCCAGTCTTC 2122
QY 983 LysIleArgAsnPheValValPheLysAsnAsnSerThrLysLysGluTyrLysLys 1002
Db 2123 AAAGTCCGCAGC-----ATGCTGCTGAGAGACCCCTCAGGGGGGATGCAAGCCAAAAG 2176
QY 1003 TTPValGluLeuProIleThrPheProAsnLeuAspTyrSerGluCysCys 1019
Db 2177 TGGTCCCGCGCTCTTCGCGTGCCTTGACTTCTTCCAGCATTC 2227
```

Search completed: May 16, 2003, 23:07:50
Job time : 4563 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:41:47 ; Search time 30 seconds
(without alignments)
1005.283 Million cell updates/sec

Title: US-09-515-363C-2
Perfect score: 5311
Sequence: 1 MSNGYSTDENFRYLSCFRA.....LPITFPNDYSECCLFSDSD 1025

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfl.es1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--|
| 1 | 194.5 | 3.7 | 2482 | 1 | US-08-328-254-6 Sequence 6, Appl |
| 2 | 191.5 | 3.6 | 3248 | 1 | US-08-353-700-1 Sequence 1, Appl |
| 3 | 191.5 | 3.6 | 3248 | 5 | PCT-US95-16216-1 Sequence 1, Appl |
| 4 | 188.5 | 3.5 | 1388 | 2 | US-08-685-576-4 Sequence 4, Appl |
| 5 | 186.5 | 3.5 | 1786 | 4 | US-08-973-462-8 Sequence 5465, Ap |
| 6 | 179 | 3.4 | 666 | 4 | US-09-134-001C-5465 Sequence 4820, Ap |
| 7 | 174 | 3.3 | 1388 | 2 | US-08-685-576-1 Sequence 1, Appl |
| 8 | 173 | 3.3 | 1211 | 4 | US-09-134-001C-4820 Sequence 4, Appl |
| 9 | 171.5 | 3.2 | 976 | 4 | US-09-104-324B-4 Sequence 2, Appl |
| 10 | 171.5 | 3.2 | 2285 | 4 | US-09-308-375-2 Sequence 2, Appl |
| 11 | 167 | 3.1 | 3696 | 4 | US-09-134-001C-5080 Sequence 3, Appl |
| 12 | 166.5 | 3.1 | 1886 | 4 | US-08-938-105-3 Sequence 2, Appl |
| 13 | 164 | 3.1 | 1354 | 3 | US-08-685-871-2 Sequence 2, Appl |
| 14 | 163 | 3.1 | 1939 | 4 | US-09-310-187A-1 Sequence 4452, Ap |
| 15 | 159.5 | 3.0 | 956 | 4 | US-09-134-001C-4452 Sequence 3, Appl |
| 16 | 158 | 3.0 | 781 | 2 | US-08-675-631-3 Sequence 4, Appl |
| 17 | 155 | 2.9 | 1261 | 4 | US-09-208-742-4 Sequence 2, Appl |
| 18 | 155 | 2.9 | 1261 | 4 | US-09-332-295-2 Sequence 2, Appl |
| 19 | 155 | 2.9 | 1261 | 4 | US-09-709-979-2 Sequence 3159, Ap |
| 20 | 154.5 | 2.9 | 10182 | 4 | US-09-134-001C-3159 Sequence 6, Appl |
| 21 | 152 | 2.9 | 407 | 3 | US-08-989-370-6 Sequence 16, Appl |
| 22 | 150 | 2.8 | 1234 | 4 | US-09-592-054-8 Sequence 8, Appl |
| 23 | 149 | 2.8 | 2329 | 3 | US-08-755-587-16 Sequence 9, Appl |
| 24 | 148.5 | 2.8 | 630 | 4 | US-08-973-462-9 Sequence 5, Appl |
| 25 | 148 | 2.8 | 407 | 3 | US-08-989-370-5 Sequence 2, Appl |
| 26 | 148 | 2.8 | 3418 | 2 | US-08-639-501-2 Sequence 2, Appl |
| 27 | 148 | 2.8 | 3418 | 3 | US-09-044-946-2 Sequence 2, Appl |

| | | | | | |
|----|-------|-----|------|---|---------------------------------------|
| 28 | 148 | 2.8 | 3418 | 3 | US-09-044-908-2 Sequence 2, Appl |
| 29 | 147.5 | 2.8 | 1461 | 2 | US-08-993-228-10 Sequence 10, Appl |
| 30 | 146.5 | 2.8 | 1098 | 4 | US-08-923-992A-8 Sequence 8, Appl |
| 31 | 146 | 2.7 | 989 | 4 | US-08-213-419B-2 Sequence 2, Appl |
| 32 | 146 | 2.7 | 989 | 4 | US-08-213-419B-2 Sequence 4, Appl |
| 33 | 145 | 2.7 | 2391 | 4 | US-08-446-855A-2 Sequence 2, Appl |
| 34 | 145 | 2.7 | 2391 | 4 | US-09-150-741-2 Sequence 2, Appl |
| 35 | 144 | 2.7 | 1388 | 4 | US-09-572-191-2 Sequence 2, Appl |
| 36 | 144 | 2.7 | 1388 | 4 | US-09-723-262-2 Sequence 2, Appl |
| 37 | 144 | 2.7 | 1388 | 4 | US-09-723-219-2 Sequence 2, Appl |
| 38 | 143 | 2.7 | 1232 | 4 | US-09-592-054-2 Sequence 2, Appl |
| 39 | 142.5 | 2.7 | 1312 | 2 | US-08-687-080-51 Sequence 51, Appl |
| 40 | 142 | 2.7 | 3418 | 2 | US-08-603-753D-1 Sequence 4, Appl |
| 41 | 142 | 2.7 | 3418 | 3 | US-08-755-587-44 Sequence 44, Appl |
| 42 | 142 | 2.7 | 3418 | 4 | US-09-099-753-4 Sequence 4, Appl |
| 43 | 142 | 2.7 | 3418 | 4 | US-08-986-106-4 Sequence 4, Appl |
| 44 | 141 | 2.7 | 1057 | 4 | US-09-341-782-10 Sequence 10, Appl |
| 45 | 141 | 2.7 | 1057 | 4 | US-09-723-820-10 Sequence 10, Appl |

ALIGNMENTS

RESULT 1
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hua
TITLE OF INVENTION: A No. 5710022zel Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6
Query Match
Best Local Similarity 19.2%: Score 194.5; DB 1; Length 2482;
Matches 219; Conservative 186; Mismatches 473; Indels 265; Gaps 46;
QY 21 RKVMYIQQEPPVL-----DYLTFLPAKVEKQIORTVATSG-----NNQAVELLIS 64
DB 1189 KIEACIELEKIVGELKENSDESKLEVFSCDHQELQRVETSEGLNSDLQEHNAQSSRE 1248

QY 65 TLEKGVML--GWTFEVALRTGSPPLAARYMPELTDLPSPFENAHIVYLOLNLQ 122
 Db 1249 DIGDNVAKVNSWKEFLD-----VENELSRIRSEKASIEHVALYLEADLE 1294
 QY 123 PTLVDKL-LVRDYLK-----CMEELTLI-EDNRILAAENNGNE--S-VHELKRIYQ 173
 Db 1295 VVOTELKLEKDNENKQKIVICLEBELSVTSERNOLRGELDTMSKKTALIQLSKME 1354
 QY 174 K-----ENMFSAFLNVLROT-----GNNELVDELGSDCSSENAELENLS-VDGPOVEK 223
 Db 1355 KTOELSHQSECLHCIOVAEAEVKEKTELLQTLIS-SDVELLKDKTHLOKIQSLKDSQ 1413
 QY 224 LLST-----VOPNLEKVMGEMNS-----SESSFADSSVSESJSTLAEGS---- 266
 Db 1414 ALSTLCGELENOJAOLNKKELVYKESBLOARLSSEDEKLVNSKALEVALVEKFEAL 1473
 QY 267 -VSCLESIGHNSMGSOGTMSGDSDE-----ENVARASPIPELOIHYOMVAP 318
 Db 1474 RLSTOEEV-HOLRQIEKLVRIEADKKOLHIAKLEKEREBSNLSKIVENLEKELQ 1532
 QY 319 ALEGNIIICLPTGSGRTVAVYIAK-DHLDKKKASEPKVIVLVNKVIVLEOLTKER 377
 Db 1533 MSENDELVIIDAKENKAVERELKQIEEMARSLKVFELDVLTIKREKENIKQIQEKQ 1592
 QY 378 QPPLAKWYVIG-----LSGDTOLKISPEVYKSCDITLSTAOIENSLIMEN----- 426
 Db 1593 Q--LSELDLILSKSLKEKEKQAEIQIKEESK-----TAVEMLONOKILNEAVALC 1644
 QY 427 GEDAGQULSDFSLIIDECHHTNKAEVYNNIMHYLMOKLKNRLKKEKVIPIYQ-- 484
 Db 1645 GDOELKATBQSLDPIIEEHOIRNST-----EKLARIADKKQICVLOOLKE 1694
 QY 485 -----LGLTASPGVGATKQAKAEHEILKI-----CANID 514
 Db 1695 SEHADLLKGRVENLERELEIARTNOEHALEENSKGVEVTLKKEGQSLRGELD 1754
 QY 515 APTIKVKNLLOLKNQIOEPCKKRIADATREDPKKELLEINTIQTYVQMSPSDBG 574
 Db 1755 VVITREKENLNELOKEGERISELEIINSSEENILOEK-----EOKRVMEKESSTA 1807
 QY 575 TOPYEGMAIOMEKKAKKKNRERYCA--EHLRYNEALQIN--CTIRHIDA-----YTHL 626
 Db 1808 MEMLOTOLKEINRYVALINDOACAKKQONLSQVECELEKALJULQIIPAKNNIYTL 1867
 QY 627 EFTFYN-----EKKOKFAVIEDSDGDEDEDDIKKPLKLETD 673
 Db 1868 QSSVNGLIQEVEDGKOKLEKDEISRLKNOI-----DDQEQVSKLSQVEGE 1915
 QY 674 RLMLTFEFNNMKRLAENPEY-----ENELTKLKLNTIMFYTRPHSARGIIFPK 726
 Db 1916 H---OLMKRONLELNLVLEOKIOVLOSKNASLODTLEVLOSSTKNLENE---LELTK 1969
 QY 727 TROSAYALQWITENEKFAEVGVKAHHLIGASHSEFKPMTOEJKEVIAKERTGKINL 786
 Db 1970 MOKMSV-----EKVAKMTAKETELORENHEMAOKTAELOFELSGSKNNLAGHLOL 2021
 QY 787 -----IATVAEBGLDKECINIVIRGLVTNEIAMVQARARADESTYV 831
 Db 2022 LBEIKSKDQLELLENSELKSLDCMHKDOVEKEKGVREELAVYOLRHHAEKHOVL 2081
 QY 832 LVAHSSGVILEHETVDFEKMMYKALHCVOMKPE-----EYAKHIFLOMOSIME 883
 Db 2082 LLDTNQYVEVLOT--YREKLTSE-ECLSQKLEIDLKSKKELNNSKATTO-TLE 2136
 QY 884 KMKMTK-----RNIAKHYKNNPSLITPLKNCISYLACSGEDIVIEK----- 925
 Db 2137 ELAKTMDLKYVNOLEKKNERAGOKMILLISCKOLF--EKKILOKHISQLOAAQK 2193
 QY 926 -----MHVNMPEPEKELYIVRENKA-----LOKKCADYUINGELITKCGOAGMTM 971
 Db 2194 OKTGVMDTKVDELTEIEIKELKETLEKKEADEVYLDKYS-LIISHKIKAKKMLELQ 2252

QY 972 MVH 974
 Db 2253 VAH 2255
 RESULT 2
 US-08-353-700-1
 : Sequence 1, Application US/08353700
 : Patent No. 5599919
 : GENERAL INFORMATION:
 : APPLICANT: YEN, TIMOTHY J.
 : TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 : TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
 : STREET: 1601 MARKET STREET, SUITE 720
 : CITY: PHILADELPHIA
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103-2307
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent in Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/353,700
 : FILING DATE: 09-DEC-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: REED, JANET E.
 : REGISTRATION NUMBER: 36,252
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (215) 563-4100
 : TELEFAX: (215) 563-4044
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3248 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHEetical: NO
 : ANTI-SENSE: NO
 : ORIGINAL SOURCE:
 : ORGANISM: HUMAN
 : US-08-353-700-1
 Query Match 3.68; Score 191.5; DB 1; Length 3248
 Best Local Similarity 19.28; Pred. No. 1.2e-06;
 Matches 221; Conservative 184; Mismatches 465; Indels 241; Gaps 47;

QY 21 RYKMTIQVEPV-----DYTLFLPAEYKQIOITYATG-----MMAVHLLS 64
 Db 1917 KIEACIELEKELKENSLSKLEFEVSCDHOELLORVETSGLSNLSDELMADSSRE 1976
 QY 65 TLEKGVML--GWTFEVALRTGSPPLAARYMPELTDLPSPFENAHIVYLOLNLQ 122
 Db 1977 DIGDNVAKVNSWKEFLD-----VENELSRIRSEKASIEHVALYLEADLE 1294
 QY 123 PTLVDKL-LVRDYLK-----CMEELTLI-EDNRILAAENNGNE--SGVRLIKRIYQ 173
 Db 2023 VVOTELKLEKDNENKQKIVICLEBELSVTSERNOLRGELDTMSKKTALIQLSKME 2082
 QY 174 K-----ENMFSAFLNVLROT-----GNNELVDELGSDCSSENAELENLS-VIRHYVEEQ 223
 Db 2083 KTOELSHQSECLHCIOVAEAEVKEKTELLQTLIS-SDVELLKDKTHLOKIQSLKDSQ 2141
 QY 224 LLST-----VOPNLEKVMGEMNS-----SESSFADSSVSESJSTLAEGS---- 266

```

Db 2142 ALSTKCELENOIAOLNKKELVKESESLQARLSESDYKILNVSQALHVAEKGFEAL 2201
QY 267 -VSCLEDSIGHNSNMGSDGTGSDSE-----ENVARASPEPELO: RPYOMEVAP 318
Db 2202 RLSTJOEEV-HOLRRIEKLVRIDEADKKOLHIAEKLKREKRENDLSK: KVENLERELO 2260
QY 319 ALEGNIIICLPTGSGKTRVAVYIAK-DHLDKKKKASEPGKVIYLVNKV: LVEQLERKEF 377
Db 2261 MSEENOLVILDAENSKAEVELTKTOIEEMARSLKIFELDLVTLRSEKLN:IKOLOEKOG 2320
QY 378 QPFLKKWYVIG-----LSGDTOLKISPEYVKSODIISTAOILENS:LNEN----- 426
Db 2321 Q--LSELDLSSFKSLLEKEQAELOIKESK-----TAVEMLONO:KLNENVAALC 2372
QY 427 GEDAGVQLSDFSIIIDECHTNKKAVYNNIRHYLMOKLNKRNK:KKNKVIPLPOT-- 484
Db 2373 GQDELKAKVQSLDPRIEEHOIRNSI-----ELKARLEADKKOLCLOOLKE 2422
QY 485 -----LGLTASPGVGAATKQAKAEHILK:-----CANLD 514
Db 2423 SEHHADLLGRVBNLELELEIARTNOEHAALEENKSGEVELKAKIEV: QSLRGLELD 2482
QY 515 APTIKTVKENLDQKNOIQEPCKKFAIADATREDPREKLEIMTR:QTYQMSPSDEG 574
Db 2483 VVIRSEKENLTLEQKEQERISELEIINSFENILOEK----- 2521
QY 575 TOPYEMALOMEKKA-----KKNRKEKRVCAEH-----LKNKICALQIN-- 614
Db 2522 ---EOKVOMKESSTAMEMLOTOLKELN--ERVAALINDQACAKEL:RSSQVECLD: 2575
QY 615 -DTRMIDAVTHLETFYNEKOKKFAVIEDSD-----BGGDQYC--TQDEDELLK 665
Db 2576 LEKAQLOGLDEKKNYIYLOSVKGLIOVEDGKOLKKEDEISHLK:IODDOLYS 2635
QY 666 PLKLDDEFLMTLFFENKMKRLAENDEY-----ENEKLTLYKNTIMEQYRTES 718
Db 2636 KLSQVGEH---OLMKEONLELNLTVLEOKIOV:QSKNASLODTLEV:SSYKNLENE 2692
QY 719 ARCIITKROSAYALSOMITENKFAEYGVANHL:IGAGHSEFPRTYONKOKEYISKP 778
Db 2693 ---LELTAKDKMSFV-----EKVNMKTAKETELOREMHEMAQTAEL:CELSGEKKN 2741
QY 779 RTGKINL-----IATTVABEGDICEKNIIVIRGLVJNE:IMVQARGRA 823
Db 2742 LAGELDLLEIKSSKQOLKELTLENSLAKSLDCMHKQOVEKSVRGE:AEYOURLHE 2801
QY 824 RADESTYVLAHSGSVIEHETVNDEREKMYKAIHQVONKPE-----EYAKHILB 875
Db 2802 AERKHQALLIDTNKOYEVEIOT---YREKLTSKR-ECLSSOKLEIDLKSSKEELNNSLK 2857
QY 876 LQMSIMEKKMKTK-----RNIAGHYKNNSPLITFLKNGSVLACSGH:IHVLEK-- 925
Db 2858 ATTO-LLEELKTKMDNLKYVNOJKRENEPAQKMLLRSKQLE--EKEKILOKELS 2913
QY 926 -----MHVNMTPPEKELYIVREKA-----LOKKAQVOYNGEILCK 963
Db 2914 QLOAAQEKOKTGYMDTKYDELTEIKELKETEKEKREADELBYKCS:ILISHKLEK 2972
QY 964 CGQAGMTMAVH 974
Db 2973 AKEMLETOVAH 2983

```

```

RESULT 3
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yeh, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYDROTHERMAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Query Match 3.6%; Score 191.5; DB 5; Length 3248;
Best Local Similarity 19.2%; Pred. No. 1.2e-06;
Matches 221; Conservative 184; Mismatches 465; Indels 291; Gaps 47;

QY 21 RYKMTYQVEPVL-----DYTLPLPAVEKQIOQVATSG-----NMQAVELLIS 64
Db 1917 KIEACIELEIKYGEKLEKNSDSEKLEVFSCDHOELLORVETSEGLNDEHHAUSRE 1976
QY 65 TLEKGYWHL--GWTREPVALRRTGSPPLAARMNELDLPSPENAHDEYLC:LNLLQ 122
Db 1977 DIGDVAKYNDSWKERFLD-----VENELSRIRSEKASIEH-EALYLEADE 2022
QY 123 PLYVDKL-LVRVDLDR-----CMEELLTI-EDRRRIAAENNGNE--SGVELKRIYQ 173
Db 2023 VVOTREKLEKLENKOKAYIVLEBELSVTISERNOLGELDITMSKKTALDQLEKMKKE 2082
QY 174 K---ENMFSAFLNVLROT-----GNNEIYQELTSGDSESNAEITLNSQ--VDGQVEEO 223
Db 2083 KTOELSHQSECLHCLQVAAEAVKKEKTELLQTLSSDVSSELLKDTHTHLOEKLOS:EKDSQ 2141
QY 224 LLSTT-----VOPNLEKEVGMENNS-----SESSPAOSSVYSESDTSLAEGS-- 266
Db 2142 ALSTKCELENOIAOLNKKELVKESESLQARLSESDYKILNVSQALHVAEKGFEAL 2201
QY 267 -VSCLEDSIGHNSNMGSDGTGSDSE-----ENVARASPEPELO: RPYOMEVAP 318
Db 2202 RLSTJOEEV-HOLRRIEKLVRIDEADKKOLHIAEKLKREKRENDLSK: KVENLERELO 2260
QY 319 ALEGNIIICLPTGSGKTRVAVYIAK-DHLDKKKKASEPGKVIYLVNKV: LVEQLERKEF 377
Db 2261 MSEENOLVILDAENSKAEVELTKTOIEEMARSLKIFELDLVTLRSEKLN:IKOLOEKOG 2320
QY 378 QPFLKKWYVIG-----LSGDTOLKISPEYVKSODIISTAOILENS:LNEN----- 426
Db 2321 Q--LSELDLSSFKSLLEKEQAELOIKESK-----TAVEMLONO:KLNENVAALC 2372
QY 427 GEDAGVQLSDFSIIIDECHTNKKAVYNNIRHYLMOKLNKRNK:KKNKVIPLPOT-- 484

```

```

Db 2373 GDOELKATBOSLDPIIEEHOIRNSI-----EILRLALEADEKKQICVLOJUNE 2422
OY 485 -----LGITASPGVGATKQAKAHENHKL-----CANLD 514
Db 2423 SEHHADLLKRVENLERELEIARTNOEHAALEAENSKGVETIKAKIEIQTOSLRELD 2482
OY 515 APTIKVKNLDOAKNOIOEPCKFAIADATREDPKREKILLEMTRIQVTVQMSPSDNG 574
Db 2483 VVIRESEKENTLNELOKEORISELEIISSENILOEK-----2521
OY 575 TOPYEQMAIOMERKAA-----KGNRRKRVCAEH-----LRKYNALQIN-----614
Db 2522 -----EOKVQOMKEKSTAMLOTOLEKLN--ERVAALINDOACKAKIUNISSQVECH 2575
OY 615 -DTIRMDAYTHLETFYNEKOKKFAVIEDSD-----JGSDXEC-----GDEDEDLKK 665
Db 2576 LEMAOQLLOGLDEAKNNYIYLOS SVKGLIOEVEDGKOKLKKDEISRLKNOIODQHOLVS 2635
OY 666 PLKLDTDGFLMTLFFENKMLKRLAENPEY-----INEXITKLRNTIMEQYTRIEHS 718
Db 2636 KLSQVEGEN-----OLMKEQUNLELNLTVLEOKIYOVLQSNASIODTELEVLOSSTYKLEML 2692
OY 719 ARGIITKTROSAYALSONITENEKFAEYGVAKAHILIGASHSSEKPMPLNEOKEVISKF 778
Db 2693 -----LEITKMDKMSFV-----EKYNNKTAKETELORIKHEMAQKATARIOBELSGEKMR 2741
OY 779 KTGKINLL-----IATVAEGLDKECNIVIRGLVTVN-AMVOARSHA 823
Db 2742 LAGELLOLLLEIKSSKQOLKELELLENSEILKSLDCMHKIDJVEKGVRIEIAEYOURLHE 2801
OY 824 RADESTEYVVAHSGSVIEHETVNDPREKMYKAIHCVNNMPE-----EYAKHILF 875
Db 2802 AEKKHQAALLDNTKQYEVLEIQT-----YREKLTSE--ECLSSQKIEIDLKSSKEELNLSK 2857
OY 876 LOMOSIMERKMTK-----RNIAKHYNKNSLITFLKNCVSILACSNDIHVLEK-----925
Db 2858 ATTO-ILFELKTKKMDNLKYVNOKKENERRAQQKMLL-KSKROLE-----KKEITLOKHLS 2913
OY 926 -----MHVNMTPFEKELYIVRENKA-----LOKCA VOINBETICK 963
Db 2914 OLQAAOEkOKGTVMQKVDLITELIKELKETLEETKIDAEVLDKYS-ILISHKLEK 2972
OY 964 CGOAGMTMNVH 974
Db 2973 AKEMLETOVAH 2983

RESULT 4
US-08-685-576-4
Sequence 4, Application US/08685576
Patent No. 5906819
GENERAL INFORMATION:
APPLICANT: Kalbuchl, Kozo
APPLICANT: Iwamatsu, Akihiro
APPLICANT: Nakano, Takeshi
APPLICANT: Ito, Masaaki
APPLICANT: Takahashi, No. 5906819uaki
TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,576

```

```

FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-325129
FILING DATE: 20-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-17150
FILING DATE: 05-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-131206
FILING DATE: 26-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/843
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-576-4

Query Match 3.58; Score 188.5; DB 2; Length 1348;
Best Local Similarity 18.2%; Pred. No. 5.3e-07;
Matches 204; Conservative 195; Mismatches 399; Indels 321; Gaps 57;

OY 71 WHLGTREPEALRRTGSPLAARYNPPL-TDLSPSEMAHDEYLOLLNLQI ILVDKL 129
Db 359 WH-W-----DNIRETAAPV-----PELSSDIDSSNFDIEDD-----KQGVETPIPIKA 402
OY 130 LYROYLD-----KMEELITIEDRRIRIAAENNGESVRELLRIYQKEMWFAFLNLV 185
Db 403 FVGNOLPIGTYIRENLILSDSP--SCRENDISQSKNEESQIDQK-----LITL 452
OY 186 ROTGNNEI-VOELTSGDSCSESNAEIEENLSQYDGPVEBOLLSTVOPALEKVMEMENS 244
Db 453 EHLSTENMOAKBELQOKKSVNTRLEKTA-----ELEELI--ILKRSVFALNOLEREK 505
OY 245 SESSFADSSVVSSEDTSL-----AEGSVSCIDESTL-----GHNSNMGSDS-GIMGSDSD 292
Db 506 ALLOHKNAEYORKADHEADKRNLENVDVNSLKDLEDLKKRNSQSISTEKVNOLOROL 565
OY 293 EENVARASPEPEIOLRYOMEVQPA-----LECKNIILCPTSGKTVAVYIA 343
Db 566 ETNALLRRESPTAARLRTQAESSKOIOLESNNRDLODKNCL--LETAKLLEKEPINL 623
OY 344 KDHLDKKKKASEPGKVILVANKVLLVEOLFPRKEFOPFLKMYRVIGLSJDTOLKISPEEV 403
Db 624 OSALSERRDRTHGSEIL-----NDLQG-----RICGLEED-----654
OY 404 VKSCDIITIAQI-----LENSLNLNENGEDAGVOLS-DESLIIT-----DECHH-----447
Db 655 LKNGILLAKVELERQOEFTDLEK-EKSNMEIDMTYOLKVIQSLSECEAEHAKTKA 713
OY 448 --TKNEAVYNNIMRHYLMQKLKNRLEKKNPVIRPQIIDLTSAPVGATKAKAEH 505
Db 714 RLADKNKIYESI-----EKAESAKMEKMKLE-----ERTLKQKVENL 753
OY 506 IL--KLCANDAPTIKTVKENLQKNOIOEPCKFAIADATREDPKREKILFIMTRIO 562
Db 754 LLEAKKCSLIDC-----DLKQSOQKITELLKQ--KDYLNED-VRNLTILIBETOQ 801
OY 563 TYQOMSPMSDFGTQPYEQMAIOMERKAAKGNRRERYCAEHLRKYNEAL-----QI 613
Db 802 KRC--LTQNDLKMQTOQVNTLKMSEKQOLKQNNHLMEMKKNLEKQNNALREKQIADQGM 859
OY 614 NDTIRMDAYTHLETFYV-----EKKOK-----KFAVIEDSD-----646

```

Db 860 KELDOLAEFOYFLYKTOVELKECEKEKTKLGELOOKKALODER:STAOLE:TL 919
 QY 647 -----EGDDDEYCGDDED-----EDDLKKPL-----KLHMDRFLV LFFENKKL 687
 Db 920 TKAOSQALARSIAEBOYSLEKEKIMKELEKEMARHKOELIKKATITSL-ETNRTL 978
 QY 688 KRLAENPEYENKELTKLRTIMEOYTRTEE-----SARGITFKROSAYA.SOMITEN-- 741
 Db 979 TSDVAILANEKELEKLNKLDOVOQLSRKLDELISAAAT-----KAOFKOLLTERITK 1031
 QY 742 -----EKFAEVYKAHHLIAGSHSEFPKPTONDOKEVISKRIJKNL:----- 786
 Db 1032 TQAVNLALAI-MNRKEPVKRGMDTVR-RKEKENRKLHMLKSRKELTLMIKYOKELN 1089
 QY 787 -LATVAEGDLKECNIYIRGYLTNEIAMOARGARA-----DESTVYVAHSGSV 840
 Db 1090 EMOAOIAEE-----SOTRIELOMTLOSKDSIDRLRSQLOALHI IDSSSI-----GSGP 1139
 QY 841 IEHETVNDPREKMM-YKAHCVONKKPREYAKHILELOMOSIMEKKMKIKRNIANKYKN 898
 Db 1140 GDAEADDPRESLBEGLSLPVANNKKKRGVKKYIVYSSKKLIFYDSLKE-----QS 1194
 QY 899 NPSLTLFLCKNSVLACSEDIHVLEKMHVANTPEFKEL--YIVRENKALOKKADY 954
 Db 1195 NPYMVLIDIKLEHVRPVOTDVR-----RADAKEIPRIHITLVANHESKEKEDEF 1244
 QY 955 QIN-GE-----IICKGQANGTMVH----- 974
 Db 1245 PVEPVEKSNYICKHCHETPTLYHPTNCEACMKPLMHMKIPALBCHIKHCKDH 1304
 QY 975 -----KGLDEPLKIRNFVYVFN-----NSTKROYKVV 1004
 Db 1305 MCKKEELIAPC-KVYYDISTAKNLLILANSTEBQ-QKW 1311
 RESULT 5
 US-08-973-462-8
 : Sequence 8, Application US/08973462B
 : Patent No. 6191270
 : GENERAL INFORMATION:
 : APPLICANT: DUBOIS, PIERRE
 : TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAIN: POLYPEPTIDE MOLECULES
 : FILE REFERENCE: 0660-0125-0 PCT
 : CURRENT APPLICATION NUMBER: US/08/973,462B
 : EARLIER FILING DATE: 1998-02-06
 : EARLIER APPLICATION NUMBER: PCT/FR96/00894
 : EARLIER FILING DATE: 1996-06-12
 : EARLIER APPLICATION NUMBER: FR 95/07007
 : NUMBER OF SEQ ID NOS: 29
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 8
 : LENGTH: 1786
 : TYPE: PRT
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
 US-08-973-462-8
 Query Match 3.5%; Score 186.5; DB 4; Length 1786;
 Best local similarity 18.8%; Pred. No. 1.2e-06;
 Matches 215; Conservative 180; Mismatches 390; Indels 359; Gaps 50;
 QY 42 VKQDIORTVATSGNMAVELLSTLEKGVHLCMTREFYEAHKKTSPLAKRYMPELTD 101
 Db 323 VEESSVAENVEESVAENVEEIVAPTVEIV-----APVLEIVADSVYE 365
 QY 102 LPSPEFMADELYO-----LNMLLP 123
 Db 366 SVAPSVESVEEVEESVAENVEESVAENVEESVAENVEESVAENVEEIVAP 425
 QY 124 TL-----VDKLLVVDLCKMEELITIEDNRILAAENNNESGVREILKRIYOKEN 176

Db 426 TVEEIVAPTVEEIVAPSVESVAPSVESVEEVEENVEESVAEN--VEESVAENVEEIVAPSV 483
 QY 177 WFSAPLANLROTKNNELVOELTGSJSSNAIENLSOYDGOPOVEOLLST-- 227
 Db 484 EESVAENVEESVAEN--VEESVAENVEESVA--ENVEEIVAPTVEEIVAPTVEEIVAPSV 539
 QY 228 --TVOPLKEKVMGNMNNSSSSFPADSSVPSQTSLSAGSVSCDESLGNSNNSQSG 285
 Db 540 VESVAPSVESV--EENVEE-----SVAENVEESVAEN--VEESVAENV-- 579
 QY 286 TWGSDSDENVAARASPEPELOLRPYOMEPVAPALEGANIITLPT--GSKTRVAVVIA 343
 Db 580 --EESVAENVEEIVAPTVEEIVAPTVEEIVAPSV--VESVAPSVESVEEVEENVEESVA 632
 QY 344 KDLHKKKASPEGVYIYLVKVLVBOLEFKREPOPLKKRYGLSGDQOLKISPEV 403
 Db 633 E--NVEESVAENVEESVAEN--VEEIVAPTVEEIV-----APVLEIVAPSV 675
 QY 404 VKSC-----DIITSAOILENSLNLNGEDAGVQSLDFSLI----- 440
 Db 676 VESVAPSVESVEEVEENVEESVAENVEESV--AENVEESVAENVEESVAENVEEIVAPTVEE 733
 QY 441 -----IIDCHHTNKEAVYNNI-----MRHYLMOKLNNRLKKNRPYP 480
 Db 734 IYAPTVEEIVAPSVESVAPSVESVEEVEENVEESVAENVEESVAENVEESVAENVEEIVAP 793
 QY 481 -LPOLLGLTASPGVGAATKOKAEHILKLCANL-----DAF--TIKTYKE 523
 Db 794 TVEEIVAPSVESVAPSVESVAENVAATNLNLLGLIETPEIDSLINIEEYKE 853
 QY 524 N-----LDLQKQIOEPCKKFA--TADATRE-----DPREKLEL-----MIRIOTY 564
 Db 854 NVYTTILENVEETTBESVTTSSNILEEIOENTITMDTIEKLEELHENVLSALENTOSE 913
 QY 565 COMSPSPDFGTOPYEOMAI-----OMEKKAKK-----GNRKEV-----CAEHLRK 606
 Db 914 EEKKEIVDIEVEEKEVATTLIEVEQAEEKSANTITELFELLENVAESNENVAENLEK 973
 QY 607 VNEALQINDTIMIDATYHL--ETFYNEBKOKKFAVIEDOSGDDDEYCDG----- 656
 Db 974 LNEYA-FNIVLDKVEEIVEISEGSLNNEMDKAFSEIFDNVKGIOENLITGM-PSIETS 1032
 QY 657 -----DEDEDLKKPLKIDETORFLMTLPENKMKLRLAENPEYENE-KLTKLKTIMEQ 711
 Db 1033 IYIOSEKAYDL-----NENYVSSILDNIENKKBGLINKLINI----- 1069
 QY 712 YRTTESARGIIFTKROSAYALSOMITENEFKFAVGYKA--HHLIG-----AGHSSEF-- 763
 Db 1070 -----STEGVQDETJTEH-----VEQNYVVDVDPAMKDOPLGILINAGG-KEMFF 1115
 QY 764 -----KPTONEOKEVISKRTGKINLLITTAABGLIKKECNI 803
 Db 1116 NLEDVFKSESIVITVEEIKDEBPQKEVEKYSIIIEEMEN--IYDVLEEKEELCLD-- 1170
 QY 804 VIRGYLTNEIAMOARGARADESTVYVAHSGSVIEHETVNDPREKMMKALHCVON 863
 Db 1171 -----KMLDA-----VEESIEI-----SSSKETTESIKO-KEKQVSI VEEVQD 1209
 QY 864 MKPEYAKHILELOMOSIMEKKMK-----TKRNIANKNNPSL 902
 Db 1210 NMDSEVEKVELL--KNMEELMKDAVEINDITSKLIEETQGLNEVDLIDMKIKMEL 1267
 QY 903 TIFLCKNSVLACSGEDI--HYIEKMHVNMTP-PEFKELYIRENKALOKKATCYDINGE 959
 Db 1268 EKALSSKEIIDAKDDTLEKVIIEEHDTITLDEVELKDVEDKI--EKVST-KDEE 1325
 QY 960 IICK 963
 Db 1326 DILK 1329
 RESULT 6

US-09-134-001C-5465
 : Sequence 5465, Application US/09134001C
 : Patent No. 6380370
 : GENERAL INFORMATION:
 : APPLICANT: Lynn Doucelte-Stamm et al
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: GTC-007
 : CURRENT APPLICATION NUMBER: US/09/134,001C
 : CURRENT FILING DATE: 1998-08-13
 : PRIOR APPLICATION NUMBER: US 60/064,964
 : PRIOR FILING DATE: 1997-11-08
 : PRIOR APPLICATION NUMBER: US 60/055,779
 : PRIOR FILING DATE: 1997-08-14
 : NUMBER OF SEQ ID NOS: 5674
 : SEQ ID NO 5465
 : LENGTH: 666
 : TYPE: PRN
 : ORGANISM: Staphylococcus epidermidis
 : US-09-134-001C-5465

Query Match 3.4%; Score 179; DB 4; Length 666;
 Best Local Similarity 19.5%; Pred. No. 1,1e-06;
 Matches 137; Conservative 114; Mismatches 230; Indels 220; Gaps 35;

QY 321 ECKNIIICL-PTGSGKTRAVYIANDHDKKKKASEPK-VIVLNKVLVQLFR--KE 376
 DB 37 ECKRHOTLLGATGCTGKTFMSVNIK-----EVSKPTLLIHNKRLAGOLVSEFK 86
 QY 377 FOPPLKKMYRVIGLSDGDDTLKISPEV-VKSDIIISTQIILNLSLNF--NEEDAGVOLS 435
 DB 87 FEPENVEYFV-----SYDYDYOPEAYVSTDTFEKDCASINDEIDLLHSATSSLEPK 140
 QY 436 DESLIT--IDECHTNKEAVYNNIM-----RHYLMQKL-----KNN----- 469
 DB 141 DQVILIASVSCIYGLGNPREYKNLVSVAVGEMERSELRKLVQVVSFNIDRQRTF 200
 QY 470 RLKKEKRPVPLPOLIGLTASPGVGATQAKAEHIKLCANLDAF--T:KTVENILQ 527
 DB 201 RVRGDVEYFV-----ASRE--MCIRVEFFGDF:DRIRF-VAY 236
 QY 528 LKNOIEPCKKFAIDA-----TREDPFK-----EKILE----- 556
 DB 237 LIGEVIREHEHTTTPASFVTRREKMKVAIRIEKELEERKELRDENKILAEORLEK 296
 QY 557 -----IMRIQTYCOMSPMSDGTQPY-----BQMAIOMK-----KA 589
 DB 297 TNYDLEMMKMGFCGIEYVHLRLRPGSTPYTLIDYFGDMLVMIDHSHVTLPOLIK 356
 QY 590 AKKGRN-KERVCAEHLRKNEALQINDTIRMDATHTLETFNEERKDKKFAVIDSDHG 648
 DB 357 MYNGDRARQVLDHGFRLPSALD-NRPLK-----FREF--EKKTKOIVVY--SATP 403
 QY 649 GDEYCDGDEDEDLKKPLIDETDRFLMTLFFENKMKLKLAEINPEVNF:KLIRNT 708
 DB 404 G-----PELEHTIDEMVQIIRPTGLLPKIDVRPT-ENK-----IDDL 442
 QY 709 MEQYRTTESARGIITFKTROSAYALSOMITENKFAEVGVAAHLLGASHSEKPKMTQ 768
 DB 443 SEIODRVDDEVLVTLTKKMSDLTYM-----KEAGIVANLV-----HSEIKTL- 489
 QY 769 NQOKEVYSFRFGKINLLIATVAEGLDIKECNIVY-----RYGLVNF:AMVQARQH- 822
 DB 490 -EKLIEIRDLRMGTDAIVGILNLRGIDIDPEVSLVILDAQKESFLRSKSLIOTIGRA 548
 QY 823 ARADESTYVL-----VAHSGSGVLEHETVNF:FRKKMYKAH 859
 DB 549 ARNDGVEVIMYADKTDSDQVYAIDETQRRREIOLAHNKEHGHTTPTIN-----KKIH 600
 QY 860 CVQNM---KPEYAHKILELOMOSIMEKKMKTKRNIKHYK 897
 DB 601 DIVISATVESDETNOQOQTELPKMKTKKEROKITENIEKEMK 641

RESULT 7
 US-08-685-576-1
 : Sequence 1, Application US/08685576
 : Patent No. 5906819
 : GENERAL INFORMATION:
 : APPLICANT: Kalbuchi, KOZO
 : APPLICANT: Iwamatsu, Akihito
 : APPLICANT: Nakano, Takeshi
 : APPLICANT: Ito, Masaki
 : APPLICANT: Takahashi, No. 5906819nak
 : TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 : NUMBER OF SEQUENCES: 16
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/685,576
 : FILING DATE: 24-JUL-1996
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 7-325129
 : FILING DATE: 20-NOV-1995
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-17150
 : FILING DATE: 05-JAN-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-131206
 : FILING DATE: 26-APR-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Bent, Stephen A.
 : REGISTRATION NUMBER: 29,768
 : REFERENCE/DOCKET NUMBER: 16887/843
 : TELEPHONE: (202)672-5300
 : TELEFAX: (202)672-5399
 : TELE: 904136
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1388 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-685-576-1

Query Match 3.3%; Score 174; DB 2; Length 1388;
 Best Local Similarity 18.5%; Pred. No. 9.5e-06;
 Matches 207; Conservative 186; Mismatches 390; Indels 334; Gaps 58;

QY 83 LRTSPPLAARYMNPETLTDLPSPSFENAHDEYQIILNLOPTLV-----D 127
 DB 364 IRETAAPV-----PELSSDIDSSNFDDIEDKDVETFFIPKAFVGNOLPFIQTYRE 418
 QY 128 KLVADVLDKMEBELLTIEDNRRIAAENNGNESGVREILKRIVQKEMFSATLVLRQ 187
 DB 419 NULLSD-SPSCKEND-SIQSRK-----NEES-QEIOKKLYTLEHLSLEIO---- 461
 QY 188 TGNNEVLVELTSGDSESNAELENISQYDGPQVEQOLSTVQVNLKEVYGMENSSSES 247
 DB 462 -AKELEOK-----CKSVNTRLEKVAK-----ELEEEI---TLRNVSTTLRQLEKALL 508
 QY 248 SFAEDSVVSEQDTSL-----AEGSVCLDESL-----GHSNNGSGDS-GTMSGINSPEEN 295
 DB 509 QHKANEYORKADHEADKRNLENDVNSLKQDLEDLKRNQNSOISTEKVYNQLOLQDDETN 568

```

QY 296 VAARASPELOLPYOMEAOPA-----LEGNIIITLPGSGKIVAVYIANDH 346
Db 569 ALLRTESDPAARLKTQOESSKOIOLESNNRDLODKN---CL-----LETKAK 615
QY 347 LDKKKKASPEGVIVLVNKLVEOLFRRKQFPLKKW-RVIGISGDTQIKIS-PPRYVK 405
Db 616 LEKE-----FINLSVLEBRDRTHGSEII-----NDLQRTSGLEEDVK 656
QY 406 SCDDIITTAQT---LENSLNLNGEDAGVOLS-DEFSIIIT---D-THH-----447
Db 657 NCKIILAKLEKROLOEFDTLEK-EKNMEIDMTYQLKVSLEDEF-EHKATKAL 715
QY 448 TKKFAVNNIMRHLYMOKKNNRKKENKRPVLPOLIGTANVGVGAF-AKAEHIL 507
Db 716 AKNNKTYEST-----EKKSEAKMEKILSE-----ERIKKOVENLIL 755
QY 508 ---KLCANIDAFITIKVENLDOLKNLOEPCKFAIADTRRPPKKEKLEIMTRIQTY 564
Db 756 EAKKRSILDC-----DLKQSOOKINELLQ---KIVLNEO-VRNLIKIEDETOKR 803
QY 565 COMSPSDFGTOPYEOMAIOMEKKAAGKNNKREKRYCAEHLRYNEAL-----QIND 615
Db 804 C-LTONDKMOTOQVNTLKMSEKOLKORNNHLLKMKSLKONALRKPHODADGOMKE 861
QY 616 TIRMDIATNLETFYN-----EKKDK-----KFAVIEDSD-----646
Db 862 LODQLAEQVSTLYTQVRELKEBCESEKTLCKELQOQKOELODBRDS-AOLEITLTK 921
QY 647 -----BGDDEYCDGDED-----EDDLKRPD---KLDETPELMTIPENNMEKR 689
Db 922 ADSEQLASIAEBOYSDELEKIMLELEKEMARUKOFLERKATIASI-EETNITLIS 980
QY 690 LBNRYEUEKILKLTNTIMEOTYTEE--SARGIITFKTOSAVALSITEN-----741
Db 981 DVANLANEKEELNNKLEAOEOLSRKDELEISAAI-----KQOFKELTERTLKTQ 1033
QY 742 --EKFAVGVKANHILGACHSSEFPKPTONEKEVYSKTRCKINL-----I 787
Db 1034 ANKLALET-MNKRERYKRNNDTVR-RKKEENKILMEIKSEREKITQOY-KYOKLINM 1091
QY 788 ATTVAEGDIEKCNIVIRYGLVTNEIAMVOAGBARA-----DESTYUIVANGSGVTE 842
Db 1092 QAOIABE---SQIRLEQMLTDSKSDIEQLRSQOALHIGLSSSI-----GSGPDD 1141
QY 843 HEVYNDFREKMM--YKATCYONMKREYANHLILELOKOSIMEKKKTKNIAKATKNNP 900
Db 1142 TEADDFEPRESLEGWLSLVRNNTKFGVKKYVIVSSKILFYUSEQDKI-----QSNP 1196
QY 901 SLITFLCKKCSYLACSGEDIHYTEKHNHNMTPREKEL---YIVREKNOJKKADYQI 956
Db 1197 YAVLIDKLFHVRPVTOYU-----RADAKEIIRITQIYANEJSSKEDEFPV 1246
QY 957 N-GE---IICKGQAGMTMNVH-----974
Db 1247 ERYGSEKSNICHHGFIFLTVHFTNCEAKMPLNMKRPVLCBRR-IKCHKCHMD 1306
QY 975 --KGLDPLCLIRNFYVVERKN-----NSTKKOYKKVY 904
Db 1307 KKEEIIAPC-KVYYDISAKNILLLANSTEEQ-OKVY 341

```

```

; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; LENGTH: 1211
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820
Query Match 3.3%; Score 173; DB 4; Length 1211;
Best Local Similarity 18.8%; Pred. No. 9,36-06;
Matches 210; Conservative 168; Mismatches 343; Indels 48; Gaps 52;

QY 44 EOIQRTVATSGNMAOAVELLSTLEKGVNHLGWTREYALRRGSPPLAARYMNIITDLP 103
Db 199 ESIOKLDHTENLNVEDILYDEGR-----VEPLKEEA-----233
QY 104 SPSPFNADDEVLOLNLQPTLVKLVLDVLDKCMEEELLTIEDNRRIA-A-NGNE 161
Db 234 -----IAKETKOLSKEMEOS--DVIVTSDIDHTYTEMQRLDERNLKSOQA-KBQO 285
QY 162 SGVRELKRI-----YOKENM-----FSAPLNVLRQTNNELVSELGTS 200
Db 286 AQINOLLQKYGKROONDYDIEKLYELVATENEYOLSGLVNLEERKNO-----337
QY 201 DCSSENA---EIEML-SQVDPQOVE-----BOLLS--TTVOPNLEKRYGMEN-----242
Db 338 --SETNAYEEELDLWLESQOISIKNEKAQONKELIADLKNKOKOLNKEVOLESTI..YISDE 395
QY 243 -----NSSESFPDSSVVS-----ESDLSLARGVSCDSESGHNSM-SDSGT 286
Db 396 OHDELEIKSNYYTILMSQSYVNNDIRPLEHTINENAKKSRIDSRLVEFNC-KDIOQ 455
QY 287 MGSDDSEENVAARASPELOLPYOMEAOPALEGKNIILCLPTGSKTRVAVYIANDH 346
Db 456 NITQOKKEVQSKKME-KVEQNIQOLE--QOLDSKRL-----SEVENKLYOAVARY 505
QY 347 LDKKKK-----ASEFGKIVLVNKLVLVEOLFRRKQFPLKKWIVIGISDGTQIKISFP 401
Db 506 NEKLSRISDLATQOEDDYTFPFGVKNHLKAKDKELR-----GIHGAVAEIVNVP 555
QY 402 -EVVSCDIIISTAQIILESSL-NLENGEDA-----GYOLDSFSLIIOECHHT-----448
Db 556 SPMTOAIEALGAS--LOHVIYDNKEDGRQAIQYLKQGLGRATPLPLNVLOPR-VAAADI 613
QY 449 -----NKEAVYNNIMRHLY-----MOKILKN-NRLKKNKRPVLPLOI 484
Db 614 KQVARGSGOFINIASDAINVSAKYONITENLGNITLIYENLKHANELARAIYRIYTL 673
QY 485 ICLTASPG---VGATKOAKA-----EHLIKICANLDAFTTKT-----520
Db 674 EBDVNVPGSGMTGGGARFKTSLISQKDELSTWRNLEDOYOROTAFEROFECKI-JABOL 733
QY 521 -----VRE-----NLDOLKNQ---IOEPCKKPAIA--DATRECF-KETLL 555
Db 734 SEQYVSASQVYNNLKEQVHNHLELDRLKTOEAHLKHEHEFEKNDGYOJCK-KETLL 793
QY 556 E-----IMTRIQTYCOMSPSDFGTOPYEOMAIOMEKKAAGKNN-ERYCA 601
Db 794 EKONHLIEIQOOLKQLESIOERYTQLSKKGKASTHOTQO---QLHQKQSDLAIVVERIKS 850
QY 602 -----EHLKRY-----NEALQINDITIRMI-----DAYHLETFYDEK-----634
Db 851 OKOYVERLQKQSDSERQKIEVNERIKLFNSDEMGKKAFALEKLOIQQOENVR-NLNOQ 910
QY 635 -----DKKFAVIEDSDDEGDEYCGDDEDDPKKPLKDE-----TDR 674
Db 911 LSEIKQORLMEKIEINSQLOKCHODILSTENHYODIKAKQSKDLYINHAJHLMOT 970
QY 675 FLMTLFFENNMKLKLAENPEYENE-----KLTKLR-----NTIM 709

```

DB 971 YOLTV-----EYDSDETINLKKKYLTKMSIDELCPVNI:NAIEOPEELN 1020
QY 710 EGYRTTESARGIIPTKTROSAYALSOWITENEKEFAEVCVK:AHHLIGA:HSSEKPMTO 768
DB 1021 ERYTFLNEOR-----TDLEAKETLEQI:IHENDKEVEGEFKTTFAVQPH:TTVERKOLJG 1075
QY 769 NQKEY-----ISKFRITG-----INLLTA----- 788
DB 1076 GGQAEILRTEDDYLSAGVDI:IVOPPKKLOHLSLLSGGRALSAIALLPFA:LKVSAPFV 1135
QY 789 -TTVAEGEDIKECNIVRYGLVTNEIAMVQARARAPESTYVI:VAH----- 835
DB 1136 ILDEVALD--EAN-VI:RYAOLNEL-----S:ETOPFIVITRHKKTMEESONLY 1182
QY 836 ----SGSGVIEHETVNDFERKKMYKAHICVQNMKEPEVA 870
DB 1183 GVTMOESGVSKIVSVN-----LNTIDEVMEEOA 1211

RESULT 9
US-09-104-324B-4
Sequence 4, Application US/09104324B
Patent No. 6232460
GENERAL INFORMATION:
APPLICANT: T reed, Ozlem; Sahin, Ugur; Pircunbashtuh, v chael
TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
TITLE OF INVENTION: And Methods For Identifying Pathologic Markers In A Sample Of
TITLE OF INVENTION: No. 6232460mal Cells
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski LLP
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,324B
FILING DATE: 25-June-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/892,702
FILING DATE: 15-July-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6232460man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5491
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 976 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-104-324B-4

Query Match 3.2%; Score 171.5; DB 4; Length 976;
Best Local Similarity 20.5%; Pred. No. 8.9e-06;
Matches 182; Conservative 132; Mismatches 33; Indels 23; Gaps 41;
QY 136 DKCMEELITIEDNRIRIAAENGNGSGVRELKRIYQKKN:SAFLVAVLQOTGNEL-V 194
DB 39 NCTEDDL-----EFPFAKTNLSKNGENIDSDPALQVN-----FLPVLLQVNGNSDCHY 87
QY 195 QF-LTGSQCSSESNAIEENLSQV-----DGPOVEQLISTTVQVPLKEKVGOMENNSSES 247
DB 88 QEGKLDSDLENS-----EGLSRVFSKLYKEAEKIKKKKYSTVA -----ELRKQES 132

QY 248 SFADSVSVSES-----DTSLAEGSVSC-LDESIGHNSNMGSQSGTMSDSDEI NVAA- 299
DB 133 KLOENRKILIEAORKAIOELOFGENEKVSILKEEIOEN-----KDLIKENNATRH 181
QY 300 -----ASPEEIOLRPYQMEVAPALAEKNIIILCPLTGSQKTRVAAVYIAHDLCK 349
DB 182 LCNILKETCANSAEKTKYEEYERETROYMDLNNIEKMTAFQELNVAQENSEDEHF 241
QY 350 KKKASEPKVIVLVKVVLLVEOLFREFKPOPLKKWYRIGLSCTOLKISPEVVKSCDI 409
DB 242 KLKED-----YKIOHLEOYKKEI-----NKEKQVSL-----L 271
QY 410 IITSAOILENSLN-----LENGEDAGVOLSPFSLI-----IIDECHITKKA -YVNN 456
DB 272 LIQITE-KENMKKDLTFLEESRDKVNOLBEEKTKLOSENLSKOSIEKOHHLTKELFDIKVS 330
QY 457 IMRHYLMOKLNNRLKKNRPVIRPOLIGLTASGVAGATKOKAKREHLIKL:VANDAF 516
DB 331 LORSYSTOKALEEDLOIATKTIQOLTE-----EKEQMESENKAKAAH--SF 375
QY 517 TI-----KTVKENLDOLKNQIOEPCKKFA-IADATREDPFKIKLEI 557
DB 376 VYTERFTVCSLEBELLRTEDQORLEKNEDQLILTMELOKSSLEBMTKILNNKIVELEE 435
QY 558 MTRIOTYQMSPMSPDGTOPYEQMAIOMEKKA-KGNRKERVCAEHLKRYNEALJIMDT 616
DB 436 LKV-----LGEKETLTYEN--KQPEKIAEELKGTEDOLILLOARE-----KEYWDL 481
QY 617 IRMDIATVHLEFPNEE-KOKKPAVIEDSDGDDGDCG-----DEDEILKKKPL 667
DB 482 EIOLAITTSQOYSKEVKDKLTELBNKLNTELTSCHNLSLENKELIOTET:NMTEL 541
QY 668 KLDETRFLMTLFEENK-----MLK-----RLAENPEYENKLTKLNNIIMEOY 712
DB 542 KNOGD-----INNKKOEERMLKQENLOETETOLRNELEVVBELKQKHQVYCKL 594
QY 713 TRTESANGI-----IFTKTROSAYALSOWITENEK--FAEYGVAAHH:IGAGHS 760
DB 595 DKSEECNNLKKOVENKKNYIELOQENKALKKGTAESKOLNYEIKVNNLEHLESK 654
QY 761 SEFKPMQNEQKEVYSKPTGKINLLATVAAEGDLKECNIYIRGLVTNE:IAMVQAR 820
DB 655 OKFGEITTYQKEIEDK-----KISENL-LEE--VEKAKVADAEVYKQKE 698
QY 821 GRARDESTYVLA-----HSGSGVIEHETVNDFERKKMYKAHICVQ----- 862
DB 699 IDKROQHIAEMVALMEKHQYDKIIEERD-----SELGLYKSKSEQDSLSRAS ETELS 754
QY 863 NKPPEEVA-KHILELOMOSIMEKKMKTKRNIAK-HYNNNSLITFL 906
DB 755 NLKAEILLSVKKQOLEIEREKELKREAKENTATLKEKKDKKTQYFL 800

RESULT 10
US-09-308-375-2
Sequence 2, Application US/09308375
Patent No. 6300117
GENERAL INFORMATION:
APPLICANT: Genencor International, Inc.
TITLE OF INVENTION: Proteases From Gram-Positive Organisms
FILE REFERENCE: GC394-PCF
CURRENT APPLICATION NUMBER: US/09/308,375
CURRENT FILING DATE: 1999-05-14
EARLIER APPLICATION NUMBER: EP9719636.4
EARLIER FILING DATE: 1997-09-15
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 2285
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 3.2%; Score 171.5; DB 4; Length 2285;
 Best Local Similarity 18.2%; Pred. No. 3.5e-05;
 Matches 214; Conservative 177; Mismatches 392; Indels 393; Gaps 50;

15 ISCFRRARVMTIQVEPVLDYTLFLPAEYKEQIQRTVATISGNMCAVELLSTLEKGVHLG 74
 353 ISGLKEMYSQAIEITLMTNI-----RRVMNPDKYINELLQFSTID---LG 395
 75 WT--REFVALRRTGSPLAARYMNPETDLPSPFENAHDELJLNLILF-----TLVD 127
 396 DTLSKNTDILQMTDGFQMGFDESELSLTITTA-----QVLNVSDILPDTYNTLTA 449

128 KLL-----VLDKMEEE-----LTIEDRNKIAAENNNESGVEILLKRIYQKE 175
 450 AMLNFIAANDSISTADKLEVDNNVAVTTLDAISIRAGSTAFPGF--LNDLI-- 504
 176 NMFSAFLNLRQTN-----NELVQ 195
 505 GYTTAIASTTRSGNIVGNSLTIFARIGNNOSIKALPQIGISVKTAGTAKKSADLIS 564
 196 ELTG-----SDGSESN-----AEIENLSQVDPQ-----VEEEL 224
 565 EVAQKMDLSDAKOKTSGVAGITQLSRPMNNMNNFSAIONAKITAN--GSAMSEQOK 624
 225 LSTTVQPNLEKEVGMENNSSESFADS-SVSESOTSLAEGSVCLDSEJCHNSHMSD 283
 625 YADSLQARYNK-----LQNNFTEPAIAASDAFTISDGIETQAA--SLTAN--SVKSVGL 680
 284 SGTMSGSD-----EENVARASPEPELQRPYQMEVAPALH--KNIITPTGSGKIV 338
 681 PLLAAVSTATLLSKNTFLAS--SLITGRAM-----DETLAASLEAGMTPA 729

339 AVYIAOHDKKKKASPEPKVIVLVKVLVLPFKERQPLKKRYRIGISGDTQLKI 398
 730 AV-----ASRVLTALRGLVSTLVGAF-----AA--GWALESLS- 765
 399 SPEPVKSCDIIISTAOILENSLLENCEADGVOLDSFLIIDECHINKEAVYNNIM 458
 766 SFPAKKKAD-----DEGOSQTNV-----EATIKNDSTIKLIO 800

459 RHYLMOKLKNR--LKKENKPVIPLPOLIGTASPGVGNATKQAAEHI--KLCANLQAF 516
 801 QYKELQKVESRSLTSDDEQYLOVYQQLAQTFPALVKYDSDQNA---LTKNKELEK- 856
 517 TIKTVENIDQLKNOIQECCKFADADREDFEKLEIM--COTYQ--HSPMSFGIO 576
 857 ALENTVEYLAKQETRDASK-----TFEDASKE--IKSKDULKOYK--ADYNKGRP 909

577 PYEOMAIQHEKKRAKKGNKREYCAE-----HLRKYNALQINOTIRMI--AVTHLEFY 630
 910 KMDLADDDDDYKAAADKAKOSMLKAOSDIESGNAKYKDSVLS--ANAYSSILSNTIKISI 969

631 NEEKDKFAVIEDSD-----DGGDE-----Y 653
 970 SDVYVK--LNLKDDLPPELEKSSSLQLEKMKKALISGRKAKRDNK--HLOSILFV 1027

654 CQDDEDEDLKLPL-----KLDETDFPLMTFFENKMLKRLAD--NP--ENEKILK-- 703
 1028 SKSDSSIDVFKMSFDKAKNINIKGDKSLSSVSEVQDLETTIADAGNEA--HGGKIKLEEL 1087

704 -----LNTIME-----QYRTESARGIIFTKRQSAVYA--QMITENKFAEYGVVA 751
 1088 DANSVDIDIAAIKESDMAQFDSVDVYLDGDFNNTKDVAPI--NDL--LKA-- 1138

1139 -----EGKSISANEANTLIDKDKELAQAIISLENGVYK--INPDEVIK--KKVKLDAYN 1188

803 IYIRYG--LVTNEL-----AMVOARGRARADE--TYLVASHSSGVLEHET 845
 1189 DMTYNSKLMKTEVNAIITLADTLRIDSLKLRERKRLDMSEAL-----SDLEVS 1242

846 VND-----FREKKMY--KATHCVQNMK-----PEYAKHILELQMSI 881

1243 INNADAKELKLEKMLQPGYSNSQIEAMOSVKSALBESYISASEATISTOHMKAOL 1302
 882 ME-----KKMKTKRNIAKHKKNNPSSLITFLCKKCSVLACSGELIHIEK 925
 1303 VEAQTSLENTMDQOEKANEERTSMYVVDKTK-----EALFK 1339

926 MH-----HVNMPPEKELY---IYENKALOK 950
 1340 VNAEIDKYNKOYNDPKYSQYKRAIKKEIKALOOK 1375

RESULT 11
 US-09-134-001C-5080
 : Sequence 5080, Application US/09134001C
 : Patent No. 6380370
 : GENERAL INFORMATION:
 : APPLICANT: Lynn Doucette-Stamm et al
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: GTC-007
 : CURRENT APPLICATION NUMBER: US/09/134,001C
 : CURRENT FILING DATE: 1998-08-13
 : PRIOR APPLICATION NUMBER: US 60/064,964
 : PRIOR FILING DATE: 1997-11-08
 : PRIOR APPLICATION NUMBER: US 60/055,779
 : PRIOR FILING DATE: 1997-08-14
 : NUMBER OF SEQ ID NOS: 5674
 : SEQ ID NO 5080
 : LENGTH: 3696
 : TYPE: PRT
 : ORGANISM: Staphylococcus epidermidis
 : US-09-134-001C-5080

Query Match 3.1%; Score 167; DB 4; Length 3696;
 Best Local Similarity 19.5%; Pred. No. 0.00018;
 Matches 200; Conservative 153; Mismatches 394; Indels 278; Gaps 44;

41 EYKQIQRTVATISGN-----MCAVELLSTLEKGVHML--QWTFEVALRRTGSPLA 92
 2641 EKANOQOSTIATHPSTIEEROEASAKLOEYVKAIAKIDGQNDVDEKTVVNAIEIE 2700

93 RYMNPELTDLPSPFENAHDEV--LQNLNLQPTLVKILVADVDKMEELIIE--- 147
 2701 NLPRTYKDKAKADYMAKEKQKLNQINSNDEATEEKLVASDN--NHVETTNCALIEDAP 2760

148 DNRRTAAEENNGESVRELLKRIYQK-----ENMFSAFLNLRQTN----- 190
 2761 DTQNVNVEKNKIGT--IRDIQPLVYKPTAKSKIESAVEKKTEINQCNATHL--FVREG 2819

191 NEL--VOELTSGDC--SESMAEIEMLSOVDGQVEQVEQLSTTVQPNLEKE-----YWG 239
 2820 NQNLQIHEKAKNDVMOQOTNOOVENAEQNSLDQI-----NNFRDFSKRNAVAEIKVA 2873

240 MNNSSSESFADSSVVSSTDSLSAEGSVCLDSEJCHNSHNSGDSGTGSDSE--NVAAK 299
 2874 QNKIKDELQEFSAQEEKDNALQH-----LDEQYKEIINSINQANT---DREYVNAKT 2924

300 ASPEPELQRPYQMEVAPALEGKNIICLPYSGKTRVAVYIANDHLKKKKA--SPGV 359
 2925 -----SGLNNTIEF--RPEYNNKKNA----- 2943

360 IYLVKVLVLEQDFKREKQPLKKRYRIGISGDTQLKI--SFPVAVKSCDIIISTAOIL 417
 2944 -----ILKLYDV-----SDTQEAIMNGYPTATD-----ELQ 2970

418 E--NSLNL--LENGDAGVQSLDSFLIIDECHHTNKKA--VYNNIMHLYLMOK--KNRL 471
 2971 EANSKLNKILDAKQOIGLA-----HTNNEVDIINYV-----SOKKKT--- 3009

472 KKENKPVIPLPOLIGTASPGVGATKQAKAEHILKLCANLDAFT-----IKTYEN 524
 3010 -----ILPRVDTKAARSVNLALAKO--LIKTFENTADVTHERDAI--NHVKEQ 3056

```

Yy 525 LDOAKNOJOECCKKFAIA-----DAREPREPKKILLIEMIKIQTCCOMP 563
Db 3057 LSLFNAITEKORDKIOAQDELFGLELMSIFINTIQTAKKA:SGMS:SOLENNINP 3116
Yy 570 MSDVGTPYEWOMIOMEKKAAGKNGKREKRYCAEHLKRYNEALQIWDITRMIDAVYHLTFE 629
Db 3117 ---YATEEPQIALNKVKAIVDDANK-----IREANTDESVAST--KRNATILLQAI 3164
Yy 630 YNEKKOKKFAVIEDDSDEGDDDEYCGDEDEDLKKPLKLUDETQKPLMTIFENKKMKIKR 689
Db 3165 SADQVQPOAFEEINMAQEIQRERINGSNDATREEKEEALQVHILVHSHITINNVK- 3223
Yy 690 LAEPYENENK-----LTKLRNTIMEQYTRTESAK:II-----IKTRIOSAVA 733
Db 3224 ---NOEYDTRDKDTEIAIHKIKISTIKQALNEITIQIDQPH:IKNNKISTYEEKSA 3280
Yy 734 LSONITENEFPAEYGVAKHHLIGAGHSSSEPKPTQNEOKQEVISAKTRGKINILATTVAE 793
Db 3281 IDKLKTAARIAEAIDKQ-----TNEEVKNIKK:SIDIRKILPLVIEIK 3425
Yy 794 EGL--DIKECNIVRYCLVY-NEIAMVQARGRADES-YLVAVI-SSG:VIEHETVQVF 849
Db 3336 SAANNEITHQAEYIR-GLINDNEATREKEDIALNOLD-TITQAVNSID:ALITNEAVKA 3484
Yy 850 RE--KNMYKAIHCYQNNKPREYAKILLET-----QMS:NEKKKK KRIYAKHYK 897
Db 3385 KEINSEININISVIAIKKPPAIA-ELQELADKKLNKFKQSQELTIFHQ:SNINELQALK 3443
Yy 898 NNPSLITFLKCKNSVYLAGCEDIHVIEKKNHVMNTPREKELYVHEKNA:QKKC---A 952
Db 3444 SAINHIQSONNESVSALLESISLIDISIE-----IAHKKIYAKAYIDGYS 3490
Yy 953 DYQIN 957
Db 3491 DDKIN 3495

RESULT 12
US-08-938-105-3
Sequence 3, Application US/08938105
Patent No. 6353151
GENERAL INFORMATION:
APPLICANT: Leinwand, Leslie A.
APPLICANT: Vikstrom, Karen L.
TITLE OF INVENTION: TRANSGENIC MODEL FOR HIVAT FALDUE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,105
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Mannel M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3595-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1886 amino acids

```

[illegible]

Best Local Similarity 19.2%; Pred. No. 0.00014;
Matches 215; Conservative 178; Mismatches 423; Indels 304; Gaps 48;

22 VVMYIOVEVLDLTLTLPVPEVKEIOIQTAVTSGNMOAVLLSLLEKGVHIGTREFVE 81
Db 832 MLYETIKIKLKLK-----SATEKEKATMKKEFGRIK-----PIT EKS-----E 869
QY 82 ALRRGSPILAAKRYMDELTDLPSPFENAHDEYLOI-----IN LQ---PVLVCKLVR 132
Db 870 ARKK-----ELEKKVSLLEQEKNDLOLOVADODINIABEERKPOLIKKIOLO 917
QY 133 DVLDCMEBELLTIEDRNTIAAENNGNSGVRELLKRI-----VAKNMSASFIN 183
Db 918 EAKVEMNERLEDEEPMNELTAKKRKLEDECSSELKDKIDDLHPLAKVFNKK--HATHN 975
QY 184 VLR-----OTGNNELOVELTSGDCSESNAEINLSQ-VKGPVPEFOLLSTVQV---NLEK 235
Db 976 KYVNLTEEMAGLDEITAKLT-----KEKKALQEHQOALMDLOVLEIDKVNLSKSKVLEQ 1031
QY 236 EVMGMNNSSESSPADSSVYS-----ESDTSLAEGSVSLDESIHNSNVSQSGTMSD 290
Db 1032 QVDDLEGSLEOEKKVRMDLERAKRKLEGLKLTQESIMULE-----NDK VLEEKIKKE 1086
QY 291 SD-----FENVARASPEPELOLRPYQMEVADPALHC-----K 323
Db 1087 POLNOONSKIEDQALALOLOKKLKENOARIELEBELAEKRAKAKVILKSDLSRELE 1146
QY 324 NIILPTSGGTRVAVVYIAKDH-----LD-----KAKKASEGKVI 360
Db 1147 EISERLEAGATSVQIEKNNKRAEFQKMRDLFEATIGHEVAAALKKKALUSVAVELG 1206
QY 361 VLNVKVLVEOLFRKEFOPFLKMKRVIGLSGDTOLKISFEVYVSCDIISTAOILINS 420
Db 1207 EOIDNLOKVKOLEKE-----KSEFKLELDVIVSMEOILAKANKLEKV 1250
QY 421 LKTLENG-----EDACVOLSDPS-----LILIFCHNHNKRAVYNNIMR 459
Db 1251 SRTLEQANEYRKLEEAORSLNDFTTQAKIQTENGELARQIE-----KCALISOLIR 1305
QY 460 HYL-----MOKLNNRLKKNRPVIRPOLILGTASPGVGATK----- 498
Db 1306 GKLSYQOMEDLK-RQLEEEGK-----AKNALAHMLJASARHDVILRQYIEE 1352
QY 499 -QAKAEHLIKCANLDAFTIKVKE-NLDOLKNOIOENKPKFA--IADA REDPREKL 554
Db 1353 TEKKELQVHLSKANSSEVQMRKTETDAIORTLEEEKAKKLAORLOQAL----- 1403
QY 555 LEIMTRIQYCGMSPMSDFGTQ-PYEQMAIOEKKAA-----KGNRKFIVACAEHLKY 607
Db 1404 -EAVEVNAKCSLEKTKRILQNEIEDLWADERSNAALVALLKIKORNTI KILAEKQVY 1462
QY 608 NEALQ-----INDTIRMIDAY---THLETFVNEEKORRFAVILIDSDGSD 651
Db 1463 EESQSELESQKAEARSLSTELFKLNAYESELHELETFKKNKNIQ---TEISDL--TE 1516
QY 652 ETCDDDEDDDLKPLKLDLTDORFLM-----TLPIYNNKMLKRLANNEYENKLT 702
Db 1517 QLGEQKVNHELEKQKQLEVEKLELOSALEAEASLEHEBECKILRAQLEHNOIKAFIER 1576
QY 703 KL--RNTIMEOYTRTESARGIIPITRQSAVALSOMITENKPAVEGVKAHHLIGAGHS 760
Db 1577 KLAKEKEEOKARRHQRVAVDSLOTSLSAETRSRNVLEKVKKMR--GDI NEMETOLSHA 1634
QY 761 SEFKPTQNEOKEVISKFPRTGKINLIATVAEGDLIEKNIVT---KVCVATNRI-- 814
Db 1635 NMAAFAQOVASLOSLSLKDTQIOLDADVAND---DLKE-NIATIVERNNNLQALHEEL 1690
QY 815 -AMVQARGAR-----ADESTYVLAHSGCVIETHTVNHREKMIT KALIHCVOMM 864
Db 1691 RAIVVEOTERSRLAEDELLETSEKRVOLLSONTSLINOKKKEKSDUTOLAEVEEVAVQEC 1750
QY 865 K-PEEVAHIL--ELQOMSIMEKKMKTKRINKAKNNNSLTLFLCKNSVILACSGEIDH 921

Db 1751 RNAEKKAKAITDAAMAAMEELKEDDTSIAHLERKKNNEOTIKDL----- 1795
QY 922 VIEKHHVMTPEFEKELYVRENKALOK-KCQADYQINCEI 960
Db 1796 ---QHRID---EAFQIALKGKKQLOKLEKARVLEBEEL 1828

RESULT 15
US-09-134-001C-4452
Sequence 4452, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NOCETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4452
LENGTH: 956
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4452

Query Match 3.0%; Score 159.5; DB 4; Length 956;
Best Local Similarity 19.2%; Pred. No. 9.3e-05;
Matches 170; Conservative 139; Mismatches 343; Indels 25; Gaps 43;

QY 252 SVVSESDTSLAEGSVCLDESLSGNSNMKSGDSGTMSGDSOEENVAARASPEPLQLRPY 311
Db 6 SNLIQIDIOSLYKFTI---DKDSSHKGNF---VRLVNNKEENVLSTIID----- 50
QY 312 OMEVAQALBEGKNITITLPTSGGTRVAVVYIAKDHDKKKKASEPGKVIYV---VKNVL 367
Db 51 QLHNCQSC---ISVARTESGLAST-----KSHFYDLKSKGVKGRITTSNYIGFNSPK 101
QY 368 LVEOLFKE-----FQPLKKKY-----RVIGLSGDDQ---LKISFPIVYVSCD 408
Db 102 MFEELKLENVEVKLTNIEGFHANGVYIEHHNHTSFITGSSLSNAKLWYEH---N 156
QY 409 IISTAOILENSLNLNLENGEDAGVOLSLSIIDECHHTKKEAV-YNNIMRHYMOKLK 467
Db 157 LFLSTHK--NDDLVN--NIKHKPDELMSDFSLTNEWINEXKOSFEVOTLOKVE--NIVVO 212
QY 468 NNRLKKNKPVIRPOLI-----GLTASPGVGC-----ATQAKAEHLIKLA-- 511
Db 213 NSEIKKFNESKLIRKNIHQEHALKSLESLRVNGEKGILISATGTGT-----LICALDV 267
QY 512 ---NIDATITIVENLDOLKNOIOEPCRRALADAT-----REDPEKEKLEIMTR 560
Db 268 RAYSDFKFLIVHNGI--LRALAEERKVPYEDSDPGLLTCKRKO-HDAKLA--PAT 322
QY 561 IQTCOMSPMSDFGTQPEQMAIOEKKAAKGNKKEVCAEHLKRYNEALQIOWITIRMI 620
Db 323 IOTLSKKNYAFNNSNHDIYIVPDEAHRTAASSVQK-----IFNFKPFLGLMIAFERT 378
QY 621 DAYTHLETFVNEKKKKFAVIEDSDG-----GDDEYCGDDEDDDLKVLKLD 671
Db 379 DELNIFELFNYNIAVEIR--LOEALLESNIIICPFHYFGVTDYIIONMSQEDAFN----- 429
QY 672 TDRFLMTIFFNNKMKLKLANPEYENKLTKLNTIMEOYTRTESARGIITIKTQSA 731
Db 430 -----LKLASNERVEH--IIKTN---YGYSGDVVKKGLIFVSSRGEA 468
QY 732 YALSQMITENKFAEVGVKAHHLIGAGHSSEFKPTQNEOKEVISKFPRTGKINIILATTV 791
Db 469 YOLANOLSKR-----GISSVGLTGKDSIA-----YRTETIOLKRFSSINIIIVDL 514

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:44:58 ; Search time 81 Seconds

(without alignments)
1666.196 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311
Sequence: 1 MSGXSTDENRYLISCRA.....LPTFPNDSYVGLFSDSD 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908170

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:*

| | |
|-----|--|
| 1: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:* |
| 2: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:* |
| 3: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:* |
| 4: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:* |
| 5: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:* |
| 6: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:* |
| 7: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:* |
| 8: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:* |
| 9: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:* |
| 10: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:* |
| 11: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:* |
| 12: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:* |
| 13: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:* |
| 14: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:* |
| 15: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:* |
| 16: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:* |
| 17: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:* |
| 18: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:* |
| 19: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:* |
| 20: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:* |
| 21: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:* |
| 22: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:* |
| 23: | /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-----------------------------|
| 1 | 5311 | 100.0 | 1025 | 22 | AAE10155 Human melanoma dif |
| 2 | 5285 | 99.5 | 1025 | 23 | AAE10155 Human RNA helicase |
| 3 | 4306 | 81.1 | 838 | 22 | AAE10165 Human RNA helicase |
| 4 | 2595.5 | 48.9 | 558 | 22 | AAU23090 Novel human enzyme |
| 5 | 1860 | 35.0 | 417 | 22 | AAU23647 Novel human enzyme |
| 6 | 1839 | 34.6 | 356 | 22 | AAE10129 Human polypeptide |
| 7 | 1769 | 33.3 | 348 | 22 | AAE10191 Human polypeptide |
| 8 | 1344 | 25.3 | 678 | 22 | AAE10108 Human protein sequ |
| 9 | 1093 | 20.6 | 304 | 22 | AAU00296 Interferon induced |
| 10 | 1093 | 20.6 | 308 | 22 | AAU00298 Interferon induced |

| | | | | | |
|----|-------|------|------|----|----------|
| 11 | 861 | 16.2 | 267 | 22 | AAU00297 |
| 12 | 861 | 16.2 | 271 | 22 | AAU00299 |
| 13 | 808 | 15.2 | 448 | 22 | AAU23654 |
| 14 | 783 | 14.7 | 166 | 22 | ABE42219 |
| 15 | 783 | 14.7 | 166 | 22 | AAE63104 |
| 16 | 783 | 14.7 | 166 | 22 | AAE75915 |
| 17 | 783 | 14.7 | 166 | 22 | AAE63026 |
| 18 | 782 | 14.7 | 154 | 22 | AAE74314 |
| 19 | 782 | 14.7 | 154 | 22 | ABE41112 |
| 20 | 666.5 | 12.5 | 357 | 21 | AAE42961 |
| 21 | 563.5 | 10.6 | 447 | 23 | ABE41171 |
| 22 | 558.5 | 10.5 | 447 | 22 | AAE95485 |
| 23 | 521 | 9.8 | 312 | 22 | AAU23099 |
| 24 | 488 | 9.2 | 92 | 22 | ABE79111 |
| 25 | 488 | 9.2 | 92 | 22 | ABE33083 |
| 26 | 488 | 9.2 | 92 | 22 | ABE18551 |
| 27 | 488 | 9.2 | 92 | 22 | AAE33880 |
| 28 | 488 | 9.2 | 92 | 22 | AAE62688 |
| 29 | 488 | 9.2 | 92 | 22 | AAE41138 |
| 30 | 488 | 9.2 | 92 | 22 | AAE65544 |
| 31 | 488 | 9.2 | 92 | 22 | AAE01876 |
| 32 | 488 | 9.2 | 92 | 23 | ABE35915 |
| 33 | 453 | 8.5 | 752 | 22 | AAE96107 |
| 34 | 406 | 7.6 | 1909 | 22 | AAE31166 |
| 35 | 395.5 | 7.4 | 764 | 22 | AAE62028 |
| 36 | 357 | 6.7 | 71 | 22 | ABE42510 |
| 37 | 357 | 6.7 | 71 | 22 | AAE76213 |
| 38 | 357 | 6.7 | 71 | 22 | AAE36323 |
| 39 | 332.5 | 6.3 | 1383 | 22 | ABE63586 |
| 40 | 332 | 6.3 | 1924 | 22 | AAE09768 |
| 41 | 324.5 | 6.1 | 190 | 22 | AAE63792 |
| 42 | 298 | 5.6 | 55 | 22 | AAE18482 |
| 43 | 298 | 5.6 | 55 | 23 | ABE40321 |
| 44 | 286 | 5.4 | 503 | 21 | AAE90944 |
| 45 | 284 | 5.3 | 502 | 21 | AAE90928 |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAE10155 | AAE10155 standard; Protein: 1025 AA. |
| ID | AAE10155 standard; Protein: 1025 AA. |
| AC | AAE10155; |
| XX | |
| DT | 29-NOV-2001 (first entry) |
| XX | |
| DE | Human melanoma differentiation associated (Mda)-5 protein. |
| XX | |
| KW | Human: melanoma differentiation associated gene; Mda-5; interferon; IFN; |
| KW | RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma; |
| KW | neuroblastoma; astrocytoma; glioblastoma; multifactor; cervical; |
| KW | breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity; |
| XX | central nervous system; cytosolic; apoptosis.. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FH | Key |
| FT | Region |
| FT | Location/Qualifiers |
| FT | 125..174 |
| FT | /note- "Reveals sequence homology with other CARD |
| FT | proteins" |
| FT | 722..823 |
| FT | /note- "Reveals significant homology to RNA helicase |
| FT | C-terminal conserved domain" |
| FT | 331..336 |
| FT | /note- "ATPase A motif" |
| FT | 443..446 |
| FT | /note- "ATPase B motif" |
| PN | WO200164707-A1. |
| XX | |
| PD | 07-SEP-2001. |

Interferon induced
Interferon induced
Novel human enzyme
Peptide #9725 enco
Human brain expres
Human bone marrow
Peptide #10063 enc
Human colon cancer
Human ovarian anti
Human OREF ORF2745
Human ovarian anti
Human protein sequ
Novel human enzyme
Human peptide #562
Peptide #589 encod
Protein #550 encod
Human brain expres
Human bone marrow
Peptide #572 encod
Peptide #581 encod
Peptide #558 encod
Human peptide enco
Putative P. abyssl
Amino acid sequenc
Recombinant P. fur
Peptide #10016 enc
Human bone marrow
Peptide #10360 enc
Drosophila melanog
Human dicer protei
Human immune/haema
Peptide #4916 enco
Human peptide enco
Cenarchaeum symbio

XX 28-FEB-2001: 2001WO-US06960.
 XX
 XX 29-FEB-2000: 2000US-051363.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 XX Fisher PB, Kang D, GopalKrishnan RV;
 XX WPI: 2001-565494/63.
 DR N-PSDB: AAD17203.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity
 XX
 XX Claim 21: Page 18-19; 152pp: English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human mda-5 protein.
 XX
 SQ Sequence 1025 AA:
 Query Match 100.0%; Score 5311; Dh 22; Length 1025;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1025; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 LPQILGLTASPGVGATQAQAAEENILKLCANLDAFTIKTVKENLIDKNOIOEPCKKA 540
 OY 541 IADATREDPFKEKLEIMTRIQTQYCOMSPMSDFGTQPEYQMAIQMEKKAAGKNGKEV 600
 Db 541 IADATREDPFKEKLEIMTRIQTQYCOMSPMSDFGTQPEYQMAIQMEKKAAGKNGKEV 600
 OY 601 AEHLKRYNEALQINDTIRMIDAYTHLETFYNEBKDKRAVIEDSDGDEYCDGDE 660
 Db 601 AEHLKRYNEALQINDTIRMIDAYTHLETFYNEBKDKRAVIEDSDGDEYCDGDE 660
 OY 661 DDLKRPKLDETDRLMTLFFENNMLRLAENPEYENKLTKLNTIMEQYTRTESAR 720
 Db 661 DDLKRPKLDETDRLMTLFFENNMLRLAENPEYENKLTKLNTIMEQYTRTESAR 720
 OY 721 GIIFTRQSAVALSOWITENEKFAEYGVKKAHHLIGASHSEFPMTONEKEYISKRT 780
 Db 721 GIIFTRQSAVALSOWITENEKFAEYGVKKAHHLIGASHSEFPMTONEKEYISKRT 780
 OY 781 GKINLLITVAEEGLDIRECNIVIRGLVNEIAMVQARARADESTYVLVAHSGGV 840
 Db 781 GKINLLITVAEEGLDIRECNIVIRGLVNEIAMVQARARADESTYVLVAHSGGV 840
 OY 841 IEHETVNDPREKMYKAIHCYQNMKPEYAKHLELOQSIEMKKMKRIIAKHKNP 900
 Db 841 IEHETVNDPREKMYKAIHCYQNMKPEYAKHLELOQSIEMKKMKRIIAKHKNP 900
 OY 901 SLITFLCKNCSVLACSGEDIHVIEKMHVNMTPFEKELYIRENKALOKKADYQINCE 960
 Db 901 SLITFLCKNCSVLACSGEDIHVIEKMHVNMTPFEKELYIRENKALOKKADYQINCE 960
 OY 961 ICKGQAGCTMMVHNGDLPCLKIRNFYVFKNNSTKQYKQWELPTTFNRLYSECL 1020
 Db 961 ICKGQAGCTMMVHNGDLPCLKIRNFYVFKNNSTKQYKQWELPTTFNRLYSECL 1020
 OY 1021 FSDPD 1025
 Db 1021 FSDPD 1025

RESULT 2
 ID AAM47798 standard; Protein: 1025 AA.
 XX
 XX AAM47798:
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH116.
 KW Human: RH116; RNA helicase; cytosolic; virocidic; anti-HIV;
 KW immunosuppressive; immunostimulatory; antithematic; antiarthritic;
 KW antiarteriosclerotic; osteopathic; antidiabetic; hepatotropic;
 KW antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine.
 OS Homo sapiens.
 XX
 XX W0200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 XX 11-MAY-2001; 2001WO-FR01441.
 XX
 XX 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 PI Bahr G, Cocude C, Capron A;
 XX WPI: 2002-082898/1.
 DR N-PSDB: ABA04908.


```

XX New polypeptide, useful for treating and diagnosing cancer or
PT inflammation, and drug screening, comprises a human polynucleotide
PT homologous to RNA helicase
XX
XX Claim 1: page 89-93; 114pp; French.
XX
XX The present sequence is the protein sequence for human RH16. RH16 is a
CC 116kDa protein and has homology to RNA helicases (DEXH box). RH16 and
CC its coding sequence are useful for treating cancer; active or chronic
CC infections (especially by HIV or hepatitis B or C); inherited genetic
CC diseases; (auto)immune diseases (particularly rheumatism, arthritis,
CC arteriosclerosis, osteoporosis and diabetes, but many others listed) and
CC to prevent graft rejection. RH16 and its coding sequence are also useful
CC for inducing, or increasing, the immune response to a vaccine.
XX
XX Sequence 1025 AA:
SQ
Query Match 99.5%; Score 5285; Dh 23; Length 1025;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1020: Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSGNSTDENFRYLISCFRARKMYIQVEPVLDTFLPAEYKEQIQRTVATSGNMQAVE 60
DB 1 MSGYSTDENFRYLISCFRARKMYIQVEPVLDTFLPAEYKEQIQRTVATSGNMQAVE 60
QY 61 LLISTLEKGVNHLGWTREVEALRRTGSPLAARYMPELIDLPSPFENAHDEYLOLLNL 120
DB 61 LLISTLEKGVNHLGWTREVEALRRTGSPLAARYMPELIDLPSPFENAHDEYLOLLNL 120
QY 121 LQPTLVKLLVRLVDLCKBEELLITIEDNRRLAAENENSGVRLLEK:VQENMFSA 180
DB 121 LQPTLVKLLVRLVDLCKBEELLITIEDNRRLAAENENSGVRLLEK:VQENMFSA 180
QY 181 FLVNLQGTNNELVQELTSDCSSESNABEINLSQVDPQVEBOLLSTTV:PLNEKEVQM 240
DB 181 FLVNLQGTNNELVQELTSDCSSESNABEINLSQVDPQVEBOLLSTTV:PLNEKEVQM 240
QY 241 ENNSSSFADSSVSESPISLAEGSVCLDESLGNSNMKSGSGMGS:SDENVAARA 300
DB 241 ENNSSSFADSSVSESPISLAEGSVCLDESLGNSNMKSGSGMGS:SDENVAARA 300
QY 301 SPEPELQRPYQMEVAQPALEGNIIITCLPTSGKTRVAVYIAKDL:DKKRASEPGKYI 360
DB 301 SPEPELQRPYQMEVAQPALEGNIIITCLPTSGKTRVAVYIAKDL:DKKRASEPGKYI 360
QY 361 VLVNKLVLVEQLFRKEPQPLAKKWRVYIGISDTQLKISPEVYKSCDI:STAQILENS 420
DB 361 VLVNKLVLVEQLFRKEPQPLAKKWRVYIGISDTQLKISPEVYKSCDI:STAQILENS 420
QY 421 LNLNENGEAGVQLSDPSLIIIDECHTNKEAYNNIMHYLMOKLKN:GKKNKPVIR 480
DB 421 LNLNENGEAGVQLSDPSLIIIDECHTNKEAYNNIMHYLMOKLKN:GKKNKPVIR 480
QY 481 LPQILCLTASPGVGAATKQAKAEHLIKCANIDAFITIKVENLQOLKNU:DEPKKRA 540
DB 481 LPQILCLTASPGVGAATKQAKAEHLIKCANIDAFITIKVENLQOLKNU:DEPKKRA 540
QY 541 IADATREDPEKLEIMTRIQTYCOMSPMSDFGTOPYLWALOMEKKA:KRNKERYVC 600
DB 541 IADATREDPEKLEIMTRIQTYCOMSPMSDFGTOPYLWALOMEKKA:KRNKERYVC 600
QY 601 AEHLRYKNEALQINDTRIMDATHTLETPENEKOKKFAVIDEDSDGG:VECDDEDE 660
DB 601 AEHLRYKNEALQINDTRIMDATHTLETPENEKOKKFAVIDEDSDGG:VECDDEDE 660
QY 661 DDLEKPLKLEDETRFLMTLFFENNMKRLAEPEYENELTKLRMTIM:GYRTRESAR 720
DB 661 DDLEKPLKLEDETRFLMTLFFENNMKRLAEPEYENELTKLRMTIM:GYRTRESAR 720
QY 721 GILFTKRGSAVALSOMITENEKFAEYGVKAHLLIGAGINSSEKPM:TONH:QVEISKFP 780
DB 721 GILFTKRGSAVALSOMITENEKFAEYGVKAHLLIGAGINSSEKPM:TONH:QVEISKFP 780

```

```

QY 781 GKINLIATVAEGLDIKECNIVIRGLVTNELAMVQARARADESTYVA:HSISGV 840
DB 781 GKINLIATVAEGLDIKECNIVIRGLVTNELAMVQARARADESTYVA:HSISGV 840
QY 841 IEHETVNDEREKMYKAIHQVQNMKPEBYAHKILELOSMIMEKMKTKR:NIJAKYKNNP 900
DB 841 IERTVNDFEREKMYKAIHQVQNMKPEBYAHKILELOSMIMEKMKTKR:NIJAKYKNNP 900
QY 901 SLITFLCKNSVSLACSGEDIHVIEKMHVNTPEPEKELYIRENKALQ:KCAVYINDEI 960
DB 901 SLITFLCKNSVSLACSGEDIHVIEKMHVNTPEPEKELYIRENKALQ:KCAVYINDEI 960
QY 961 ICKCGAMGTMMVHKGDLPCIKIRNFVVPKNNSTKQYKKVVELP:PIFPNLDYSECL 1020
DB 961 ICKCGAMGTMMVHKGDLPCIKIRNFVVPKNNSTKQYKKVVELP:PIFPNLDYSECL 1020
QY 1021 FSDSD 1025
DB 1021 FSDSD 1025

RESULT 3
AAE10165
ID AAE10165 standard; Protein: 838 AA.
XX
XX AAE10165:
XX
XX 29-NOV-2001 (first entry)
XX
XX RNA helicase conserved motif of human Mda-5 protein.
XX
XX Human: melanoma differentiation associated gene; Mda-5; interferon: IFN:
XX RNA helicase motif; caspase recruitment domain; CARD; therapy: melanoma;
XX neuroblastoma; astrocytoma; glioblastoma; multifome; cancer: cervical;
XX breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity:
XX central nervous system; cyostatic; apoptosis.
XX
XX Homo sapiens.
XX
XX W0200164707-A1.
XX
XX 07-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US06960.
XX
XX 29-FEB-2000; 2000US-0515363.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Fisher PB, Kang D, GopalKrishnan RV;
XX WPI: 2001-56494/63.
XX
XX Nucleic acid sequences encoding a Melanoma Differentiation Associated
XX Gene useful for cancer cell growth suppression, apoptosis and
XX anti-viral activity -
XX
XX Example 1: Fig 1D: 152pp: English.
XX
XX The present invention relates to an isolated nucleic acid encoding a
XX melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
XX contains a caspase recruitment domain (CARD) and a RNA helix use motif.
XX Mda-5 is a novel interferon (IFN) inducible gene with structural
XX similarities to RNA helicases and CARD motif containing proteins. Mda-5
XX is induced during terminal differentiation in human melanoma cells
XX treated with the combination of recombinant fibroblast IFN and the
XX antileukemic compound mezerein (MEZ). Mda-5 is useful for identifying
XX compounds that may induce its expression. Mda-5 is useful for treating
XX cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
XX multifome, cervical cancer, breast cancer, colon cancer, prostate
XX cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
XX a cancer of the central nervous system and apoptosis. The Mda 5 promoter

```

CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is RNA helicase conserved motif of human Mda-5
 CC protein.
 CC
 XX
 SQ Sequence 838 AA;
 Query Match 81.1%; Score 4306; Dh 22; Length 838;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LLLSTLEKGYMLGWTREFEVALRRTGSPLAAYNPELIDLPSPSPFNADHEYLQILNI 120
 DB 1 LLLSTLEKGYMLGWTREFEVALRRTGSPLAAYNPELIDLPSPSPFNADHEYLQILNI 69
 QY 121 LQPTLVKLLVRYDVKCKMEELLTIEDNRNIAAFNN;NESCVRRLKIVOKENMPSA 180
 DB 61 LQPTLVKLLVRYDVKCKMEELLTIEDNRNIAAFNN;NESCVRRLKIVOKENMPSA 120
 QY 181 FLVNLQGTGNNELVOELTGSSENAETENLSOVGVPOVEBOLLSTTV;PNIKEKWKIM 240
 DB 121 FLVNLQGTGNNELVOELTGSSENAETENLSOVGVPOVEBOLLSTTV;PNIKEKWKIM 180
 QY 241 ENNSSSFADSSVSESOTSLAEGSVCSLDESLGHNSNMGSDSGTMS;SDEENVAARA 300
 DB 181 ENNSSSFADSSVSESOTSLAEGSVCSLDESLGHNSNMGSDSGTMS;SDEENVAARA 240
 QY 301 SPPELQRLPYOMEVAPALGKNNITICLPTGSGKTRVAVYIAKDHLDKKKASEPKVI 360
 DB 241 SPPELQRLPYOMEVAPALGKNNITICLPTGSGKTRVAVYIAKDHLDKKKASEPKVI 300
 QY 361 VLVNKKVLVBOLEFRKEFOPFLKMYRVLIGLSGDTOLKISPEPVKSCDII;STAQILNLS 420
 DB 301 VLVNKKVLVBOLEFRKEFOPFLKMYRVLIGLSGDTOLKISPEPVKSCDII;STAQILNLS 360
 QY 421 LNLNLEGEAGVOLSDFSLIIDECHHTNKEAVYNNIMKHYYLMOKKNNNI;KKENKPIV 480
 DB 361 LNLNLEGEAGVOLSDFSLIIDECHHTNKEAVYNNIMKHYYLMOKKNNNI;KKENKPIV 420
 QY 481 LPQILGLTASPGVGATKQAKAEHLKLCANIDAFIK;VKNENDOLKNOIOEPCKFA 540
 DB 421 LPQILGLTASPGVGATKQAKAEHLKLCANIDAFIK;VKNENDOLKNOIOEPCKFA 480
 QY 541 IATATREDPKKELIEIMRIOTYGGMSPMSDGTOPYMAIOMKKAAGKGRKRYC 600
 DB 481 IATATREDPKKELIEIMRIOTYGGMSPMSDGTOPYMAIOMKKAAGKGRKRYC 540
 QY 601 AEHLRYNEALQINDTIRMIDATYHLETFYNEEKDKFPAVIDEDSDGGLDEYCDGDEDE 660
 DB 541 AEHLRYNEALQINDTIRMIDATYHLETFYNEEKDKFPAVIDEDSDGGLDEYCDGDEDE 600
 QY 661 DDLKKPLKIDETRFPLMTLFEFENKMKRLAENPEYENKLT;LNTIMEOYTRTEESAR 720
 DB 601 DDLKKPLKIDETRFPLMTLFEFENKMKRLAENPEYENKLT;LNTIMEOYTRTEESAR 660
 QY 721 GIIFTRTOSAVYALSOITENEFKFAEYGVKAHHLIGAGHSSEKPTOM;OKEVISKFRT 780
 DB 661 GIIFTRTOSAVYALSOITENEFKFAEYGVKAHHLIGAGHSSEKPTOM;OKEVISKFRT 720
 QY 781 GKINLLIATTVAAEGDLKECNIVIRYGLVTNEIAMVYARGKARADESYVAV;VAHSGSGV 840
 DB 721 GKINLLIATTVAAEGDLKECNIVIRYGLVTNEIAMVYARGKARADESYVAV;VAHSGSGV 780
 QY 841 IEHEVTNDPREKMMYKAIFCYOMKPREYAHKLELOMOSIMKKKTKRNI;AKAHKN 898
 DB 781 IEHEVTNDPREKMMYKAIFCYOMKPREYAHKLELOMOSIMKKKTKRNI;AKAHKN 838

RESULT 4
 AAU23090
 ID AAU23090 standard; Protein; 558 AA.
 AC
 XX AAU23090;
 XX

DT 18-DEC-2001 (first entry)
 XX
 DE Novel human enzyme polypeptide #176.
 XX
 KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW lase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
 KW nephrotropic; anticoagulant.
 XX
 OS Homo sapiens.
 XX
 PN W0200155301-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001W0-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198125.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225216.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233406.
 PR 14-SEP-2000; 2000US-0233066.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234224.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-465566/50.
 XX N-PSDB: AAS40960.
 DR Novel polypeptides and polynucleotides useful for diagnosis,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX Claim 11; SEQ ID No 1086; 1180pp; English.
 PS The present invention relates to the isolation of novel human enzyme
 XX polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme (AAS40785-AAS41684) and genomic sequences
 CC the functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAI22915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 558 AA;
 OY Query Match 48.9%; Score 2595.5; DB 22; Length 558;
 OY Best Local Similarity 94.6%; Pred. No. 7.2e-193;
 Db Matches 510; Conservative 10; Mismatches 16; Indels 3; Gaps 2;
 OY 294 ENVARASPEPEIOLRPYOMEVADPALEGGNITICLPFGSGKTRVAVYIAKKKA 353
 Db 1 ENVARASPEPEIOLRPYOMEVADPALEGGNITICLPFGSGKTRVAVYIAKKKA 60
 OY 354 SEPQVIVLVNKKVLLVEOLFRRKEFPLKKWRYVIGLSGDLQKLTSPPEVNSC 411ST 413
 Db 61 SEPQVIVLVNKKVLLVEOLFRRKEFPLKKWRYVIGLSGDLQKLTSPPEVNSC 111ST 120
 OY 414 AQLLENSLLNLENGEDACVQSLSDPSLITIDCHHTNKCAVYNNIMRHYLKK 473
 Db 121 AQLLENSLLNLENGEDACVQSLSDPSLITIDCHHTNKCAVYNNIMRHYLKK 180
 OY 474 ENKPVIPPLPOLLGLTASPGVGATQAAKEHILKLCANLDAFTIKYKENIC 473
 Db 181 ENKPVIPPLPOLLGLTASPGVGATQAAKEHILKLCANLDAFTIKYKENIC 240
 OY 534 EPCKFAIADATREDPEFEKLEIETWTRIOTYCOMSPMSDFCTOPYEOWAIONFKA 593
 Db 241 EPCKFAIADATREDPEFEKLEIETWTRIOTYCOMSPMSDFCTOPYEOWAIONFKA 300
 OY 594 NKKEVYCAEHLKKYNEALQINDTIRMIDAYTHLETFYNEEKDKKRAVIEDDIF 653
 Db 301 NKKEVYCAEHLKKYNEALQINDTIRMIDAYTHLETFYNEEKDKKRAVIEDDIF 360

OY 654 CGDEDEDDKKPKLDETRRLMTLFFENNMKRLAIINPEYENKELTKIINTIMEQYT 713
|||||
DB 361 CGDEDEDDKKPKLDETRRLMTLFFENNMKRLAIINPEYENKELTKIINTIMEQYT 420
OY 714 RTEESARGIIFTKROSAYALSQWITENKEFAVGVKAIHLLIGASHSEFKMTONEQKE 773
|||||
DB 421 RTEESARGIIFTKROSAYALSQWITENKEFAVGVKAIHLLIGASHSEFKMTONEQKE 480
OY 774 VISKFPRTGKINLLIATTVAEGLDKECNIVIRGLVTNEIAP--VQARGARADESTYV 831
|||||
DB 481 VISKFPRTGKINLLIATTVAEGLDKECNIVIRGLVTNEIAP--VQARGARADESTYV 537

RESULT 5
AAU23647
ID AAU23647 standard; Protein; 417 AA.
XX
AC AAU23647;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #733.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO20015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232337.
PR 14-SEP-2000; 2000US-0232338.
PR 14-SEP-2000; 2000US-0232339.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

| | |
|---------------------------|---|
| XX | The invention relates to human nucleic acids (AA157798; AA161349) and |
| CC | the encoded polypeptides (AA186642-AA142213), with neurotrophic, |
| CC | immunosuppressant and cytostatic activity. The polynucleotides are useful |
| CC | in gene therapy. A composition containing a polypeptide or polynucleotide |
| CC | of the invention may be used to treat diseases of the peripheral nervous |
| CC | system, such as peripheral nervous injuries, peripheral neuropathy and |
| CC | localised neuropathies and central nervous system diseases, such as |
| CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic |
| CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the |
| CC | utilization of the activities such as: immune system suppression, |
| CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic |
| CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, |
| CC | assays for receptor activity, arthritis and inflammation, leukaemias and |
| CC | C.N.S disorders. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification. |
| XX | |
| SQ | Sequence 356 AA: |
| Query Match | 34.6%; Score 1839; DH 22; Length 356; |
| Beat Local Similarity | 99.4%; Pred. No. 2.3e-134; |
| Matches 350; Conservative | 0; Mismatches 2; Indels 0; Gaps 0; |
| OY | 674 RFLMTLFFEDNNKMLRLAENPEYENEKLTKLNPTIMEOYTRFESARGIITKTROSAYA 733 |
| Dd | 5 RFLMTLFFENNKKLRLENPEYENEKLTKLNPTIMEOYTRFESARGIITKTROSAYA 64 |
| OY | 734 LSONTTENEKFAYGVNKAHLHGAGHSSEFKPNTONEKEVSKFPATCKINLLIATTVAE 793 |
| Dd | 65 LSONTTENEKFAEVGKKAHHLGAGHSSEFKPNTONEKEVSKFPATCKINLLIATTVAE 124 |
| OY | 794 EGLDIKECNIVIRGLVTNELIAMVGARARADESTYYLVAHSGSVIEHTVNDPREKM 853 |
| Dd | 125 EGLDIKECNIVIRGLVTNELIAMVGARARADESTYYLVAHSGSVIEHTVNDPREKM 184 |
| OY | 854 MYKAHCQNMMREPYAKHKLELOMOSIMKKMKRKN'AKHYKNPNSLITLCKKCSYL 913 |
| Dd | 185 MYKAHCQNMMREPYAKHKLELOMOSIMKKMKRKN'AKHYKNPNSLITLCKKCSYL 244 |
| OY | 914 ACSGDIDHVEEMHHVNMTPEFKELYIVENKALOKKCADVOINGRIDCKDQAMOTMV 973 |
| Dd | 245 ACSGDIDHVEEMHHVNMTPEFKELYIVENKALOKKCADVOINGRIDCKDQAMOTMV 304 |
| OY | 974 HKGIDLPCILKINFVFVKNNSTKQKYKRWVLPITFPHLDSECECLFSIID 1025 |
| Dd | 305 HKGIDLPCILKINFVFVKNNSTKQKYKRWVLPITFPHLDSECECLFSIID 356 |
| RESULT 7 | |
| ID | AAAA1915 standard; Protein: 348 AA. |
| XX | AAAA1915; |
| AC | AAAA1915; |
| XX | 22-OCT-2001 (first entry) |
| Dt | |
| DE | Human polypeptide SEQ ID NO 6846. |
| XX | |
| KW | Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer; |
| KW | peripheral nervous system; neuropathy; central nervous system; CNS; |
| KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; |
| KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; |
| KW | chemokinetic; thrombolytic; drug screening; arthritis; inflammation; |
| KW | leukemia. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | MO200153312-A1. |
| PD | 26-JUL-2001. |
| XX | |
| PF | 26-DEC-2000; 2000WO-US34263. |

| | | | | |
|-----------------------|-----------------|---|-----------------|-------------|
| PR | XX | 21-JAN-2000; | 2000US-0488725. | |
| PR | XX | 25-APR-2000; | 2000US-0552317. | |
| PR | XX | 09-JUL-2000; | 2000US-0598042. | |
| PR | XX | 19-JUL-2000; | 2000US-0620312. | |
| PR | XX | 03-AUG-2000; | 2000US-0634350. | |
| PR | XX | 14-SEP-2000; | 2000US-0662191. | |
| PR | XX | 19-OCT-2000; | 2000US-0693036. | |
| PR | XX | 29-NOV-2000; | 2000US-0727344. | |
| PA | XX | (HYSE-) HYSEQ INC. | | |
| PA | XX | Tang Y.T., Liu C., Asundi V., Chen R., Ma Y., Qian X.B., Rao F., Wang D. | | |
| P1 | PI | Wang Z., Wang Z., Mehran T., Xu C., Xue A.J., Yang Y., Zhang J. | | |
| P1 | PI | Zhao Q.A., Zhou P., Goodrich R., Drmanac R.T. | | |
| XX | XX | WPI: 2001-442253/47. | | |
| DR | DR | N-PSDB; AA161071. | | |
| PT | PT | Novel nucleic acids and polypeptides, useful for treating disorders | | |
| PT | PT | such as central nervous system injuries - | | |
| PS | PS | Example 2: SEQ ID NO 6846: 10078bp: English. | | |
| XX | XX | The invention relates to human nucleic acids (AA157798-AA161859) and | | |
| CC | CC | the encoded polypeptides (AA38642-AA42213) with neurotrophic, | | |
| CC | CC | immunosuppressant and cytostatic activity. The polynucleotides are useful | | |
| CC | CC | in gene therapy. A composition containing a polypeptide or polynucleotide | | |
| CC | CC | of the invention may be used to treat diseases of the peripheral nervous | | |
| CC | CC | system, such as peripheral nervous injuries, peripheral neuropathy and | | |
| CC | CC | localised neuropathies and central nervous system diseases, such as | | |
| CC | CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic | | |
| CC | CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the | | |
| CC | CC | utilisation of the activities such as: immune system suppression, | | |
| CC | CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic | | |
| CC | CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, | | |
| CC | CC | assays for receptor activity, arthritis and inflammation, leukaemias and | | |
| CC | CC | C.N.S. disorders. | | |
| CC | CC | Note: The sequence data for this patent did not form part of the printed | | |
| CC | CC | specification. | | |
| SQ | SQ | Sequence 348 AA: | | |
| Query Match | 33.3%. | Score 1769: | DB 22: | Length 144: |
| Best Local Similarity | 97.4%. | Pred. No. 6.3e-129: | | |
| Matches 337: | Conservative 2: | Mismatches 7: | Indels 0: | Gaps 0: |
| QY | 680 | FFENNMKMLKRLAENEVEYENKLTLRNTIMEQYRTTEESARGILFTKTRQSAVVAASQWIT | 739 | |
| DB | 3 | FLKTIKXILNRLAHEHXEYENKLTLRNTIMEQYRTTEESARGILFTKTRQSAVVAASQWIT | 62 | |
| QY | 740 | ENKEAEVGVKAHHLIGAGSHSEFPRTONEOEKEYISKPRTRKINLLATTYAERLDDIK | 799 | |
| DB | 63 | ENKEAEVGVKAHHLIGAGSHSEFPRTONEOEKEYISKPRTRKINLLATTYAERLDDIK | 122 | |
| QY | 800 | ECNIVIRGLVYNEIYAMQARGRAADESTYYLVASHSGVIEHEHYVNDPFRKKMVAIIR | 859 | |
| DB | 123 | ECNIVIRGLVYNEIYAMQARGRAADESTYYLVASHSGVIEHEHYVNDPFRKKMVAIIR | 182 | |
| QY | 860 | CVONNKPPEYAAHKILELQMSIMEKMTTRKRIAAHYKNPNSLITFLCKNCVLAASGED | 919 | |
| DB | 183 | CVONNKPPEYAAHKILELQMSIMEKMTTRKRIAAHYKNPNSLITFLCKNCVLAASGED | 242 | |
| QY | 920 | IHAVLEKMHVNVTPPEFKELYIYREKKAALOKKCAVDQINGELIICKGQGMGTMMVHKIDIL | 979 | |
| DB | 243 | IHAVLEKMHVNVTPPEFKELYIYREKKAALOKKCAVDQINGELIICKGQGMGTMMVHKIDIL | 302 | |
| QY | 980 | PLKLRNFVYVVFKNKSTKKYKKVVELPTTPPNLDYSECCLFSDDD | 1025 | |
| DB | 303 | PLKLRNFVYVVFKNKSTKKYKKVVELPTTPPNLDYSECCLFSDDD | 348 | |

| Query Match | Similarity | 33.3% | Score 1769 | DB 22 | Length 144 |
|-------------|------------|--|--------------------|------------|------------|
| Best Local | Similarity | 97.4% | Pred. No. 6.3e-129 | | |
| Matches | 337 | Conservative | 2 | Mismatches | 7 |
| | | | | Indels | |
| | | | | Gaps | 0 |
| QY | 680 | FFENNMMLKRLAENEYEENEKLTLRNTIMEQYRTTEESARGILEFTKTRQSAVVAALNQWIT | 739 | | |
| Db | 3 | FLKTIKXILNRLAENHXYENEKLTKIRNTIMEQYRTTEESARGILEFTKTRQSAVVAALNQWIT | 62 | | |
| QY | 740 | ENEKFAEYGVKAHHILGAGHSSEFRPMTOEOKEYISKRPTRKINLLIATTVAEHLLDIK | 799 | | |
| Db | 63 | ENEKFAEYGVKAHHILGAGHSSEFRPMTOEOKEYISKRPTRKINLLIATTVAEHLLDIK | 122 | | |
| QY | 800 | ECNIIYIRGLVYNELIAMVQARGRAADESTYVLVAHSGVYIEHETVNDYFKIKMVAIAH | 859 | | |
| Db | 123 | ECNIIYIRGLVYNELIAMVQARGRAADESTYVLVAHSGVYIEHETVNDYFKIKMVAIAH | 182 | | |
| QY | 860 | CVQNNKPREYAAKLTLELOMOSIMEKKMPTKRMIAHYKNPNLSITFLCNCNVLAASGED | 919 | | |
| Db | 183 | CVQNNKPREYAAKLTLELOMOSIMEKKMPTKRMIAHYKNPNLSITFLCNCNVLAASGED | 242 | | |
| QY | 920 | IHAVLEKMHVNNTPPEFKELYIVREKKAALOKKADYQINCEILICCKGQAMGTIMVHKISIDL | 979 | | |
| Db | 243 | IHAVLEKMHVNNTPPEFKELYIVREKKAALOKKADYQINCEILICCKGQAMGTIMVHKISIDL | 302 | | |
| QY | 980 | PLKLIIRNFVVVYFKNNSTKKQYKKWVELDITPEPNLDYSECCLEFSDSD | 1025 | | |
| Db | 303 | PLKLIIRNFVVVYFKNNSTKKQYKKWVELDITPEPNLDYSECCLEFSDSD | 348 | | |

AAB93708
 ID AAB93708 standard; Protein; 678 AA.
 AC AAB93708;
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:13299.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 PE 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T.
 DR MPI: 2001-318749/34.
 PT primer sets for synthesizing polynucleotides; particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 13299; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAB93893 represent human amino acid sequences; and AAH1629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 678 AA:
 Query Match 25.38; Score 1344; DB 22; Length 678;
 Best Local Similarity 40.68; Pred. No. 2e-9;
 Matches 291; Conservative 136; Mismatches 238; Indels 52; Gaps 12;
 QY 306 LQRPOMEYAPALGKNNIIICLPGSGKTRAYAIADHLDKKKASFGKVIYLVNK 365
 DB 1 MELRSTOMEYIMPALGKNNIIICLPGAGKTRAAVAIVAKRHE---IVGKVVVLLVNF 56
 QY 366 VLLVEQLFKREFQFLKKWRYVIGLSGDTOLKISPEVVKSCDIISTA. LENSILNH 425

DB 57 VHLVYQ-AGEEFRRLDGRMTVTTLISGDMGPRAGCHLARCHDILLCTAELICMTISPE 115
 QY 426 NGEDAGYOLDSFSLIIIDECHTNKEAYNNIMRTYIMOKLKNKKKNNKVI 160IL 485
 DB 116 --EEHEVELVPSLIVDECHTHRDYVNNIMSOYLLEKLRLAQ-----LIPVL 164
 QY 486 GLTASPGVGATGAKAEHILKLCANLDAFTIKVKNENLQKNOIQEPCKKILATAT 545
 DB 165 GLTASPGVGATGAKAEHILKLCANLDAFTIKVKNENLQKNOIQEPCKKILATAT 224
 QY 546 REDPPEKLEIMTRIQTGYCOMSPMS-DFGTPQYDQMAIQEKKAKKNGKPKYAEHL 604
 DB 225 SDDPGDLKKLMQDINHLEMPESRKFGTQYQVVKLSAALAGLQFGKRYVALTL 284
 QY 605 KRYNALQINDIYRMIDAVTHLETYYNEKDKKFAVIEDSDGSDDEYCDLQDCLK 664
 DB 285 RRYNDALLHDYRAVDLALADDFYHREHYTKTIL-----C----- 322
 QY 665 KPLKLDETDRFLMTLEFFENNNKMLKRLAENPEYENKTLKRLTMEQYTRTESAKGIF 724
 DB 323 -----AERLLALFDCKNELHLATHGP-ENKLEKLEKILQROTSSNSRGIIIF 373
 QY 725 TKTRQSAVALSQWITENKFAEVGKAAHLIGAGHSEFKPTONEQKEVLSKIN 784
 DB 374 TRTRQSAHSLMLLQDQGLQTVDIRADLLIGAGHSSQSTHTHQDQGVQKFLGLLN 433
 QY 785 LLIATVAEESGIDIECNVIRYGLVNEIAMVQARGARADESTYLVLAHSGVIEHE 844
 DB 434 LLVATVAEEGIDIECNVIRYGLVNEIAMVQARGARADESTYLVLAHSGVIEHE 493
 QY 845 TVNDFREKMYKAIHCYONMKPEYAHKILFLOMOSIMEKKKTRNIAKHKNPSLIT 904
 DB 494 LINEALETLMEDVAVAAYQMDQAEYQAKRDLQQAALTKRAAQAQRENORCFVEVVO 553
 QY 905 FLCKKSVLACSGEDIHVLEKHHVNMTPPEKELY-IVRENKALOKKCAQYQIN 963
 DB 554 LKICNOMVAVGSGDLRKVEGTHHVNPNFENYVNSDPAIVIKVFKDKMPCGVISCR 613
 QY 964 -CGQAGMTMVKHGLDPLCLIRNFVVFKNSTKKQYKKVYELPITPNLDYSQV 1019
 DB 614 NCGEYWGLOMIVKSVKLPVLKVR--MLLETPOGRIOAKMSRVSPVDFDLQHC 668
 RESULT 9
 ID AAU00296 standard; Protein; 304 AA.
 XX
 AC AAU00296;
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN4.
 KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune mediated glomerulonephritis;
 KW haematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN4.
 OS Homo sapiens.
 PN WO200118208-A2.
 PD 15-MAR-2001.
 PE 08-SEP-2000; 2000WO-0524704.
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.

```

XX 07-SEP-2000; 2000US-0656633.
PA (CURA-) CURAGEN CORP.
XX (BioJ ) BIOGEN INC.
XX
XX
XX Payman JA, Da Silva A, Hochman P, Hsu A;
DR MPI: 2001-235201/24.
XX N-PSDB: MMS01149.
XX
PT New interferon induced polypeptides and polynucleotides useful for the
PT diagnosis, prevention and treatment of immunological, cell
PT proliferative disorders, such as lupus erythematosus, cancer, stroke
PT and Alzheimer's disease -
XX
XX Claim 1; Page 29-32; 134pp; English.
PS
XX The sequence represents interferon induced polypeptide, IFN4, IFN
CC nucleic acids and polypeptides are useful for treating or preventing a
CC pathology associated with IFN polypeptide in a human. They are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC acids, polypeptides and antibodies are useful for diagnosis, prevention
CC or treatment of variety of immunological and cell proliferative
CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC immunodeficiency diseases such as acquired immunodeficiency syndrome
CC (AIDS), graft rejection, viral infections including hepatitis and human
CC immunodeficiency virus (HIV), immune-mediated glomerular nephritis,
CC haematologic diseases such as aplastic anaemia and chronic neutropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon adjuvants, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans.
XX
SQ Sequence 304 AA;
Query Match 20.6%; Score 1093; DH 22; Length 304;
Best Local Similarity 100.0%; Pred. No. 1,8e+76;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY 41 EKKEDIQRTVATSGNNQAVELLSTLEKGVMLGTRFEVALRRIGSPFAARYNPET 100
Db 87 EFKEDIQRTVATSGNNQAVELLSTLEKGVMLGTRFEVALRRIGSPFAARYNPET 146
QY 101 DLSPSFENAHDEYLQLNLNLQPTLVDKLVLVDVKCKEPEELITIEDNRFAAEFNNGN 160
Db 147 DLPSPSFENAHDEYLQLNLNLQPTLVDKLVLVDVKCKEPEELITIEDNRFAAEFNNGN 206
QY 161 ESGVAELAKRIYOKKNMFSAPFLNVRROGMNELVOELTSDCSNAELFPHLSOYDGPV 220
Db 207 ESGVAELAKRIYOKKNMFSAPFLNVRROGMNELVOELTSDCSNAELFPHLSOYDGPV 266
QY 221 EEOLLSTVOPRLKEVWGMENNSESSESSPADSSVYS 256
Db 267 EEOLLSTVOPRLKEVWGMENNSESSESSPADSSVYS 302
RESULT 10
ID AAU00298 standard; Protein: 308 AA.
XX AAU00298;
XX AC AAU00298;
XX 12-SEP-2001 (first entry)
XX Interferon induced polypeptide, IFN6.
XX
```

| DB | Query Match | Best Local Similarity | Matches | 216: Conservative | 0: Mismatches | 0: Indels | 0: Gaps | 0: Indels | 0: Gaps |
|----|-------------|---|---|--------------------|---------------|-------------|---------|-----------|---------|
| XX | Sequence | 308 AA: | 20.6%; | Score 1093; | DB 22: | Length 403: | | | |
| XX | Query Match | Best Local Similarity | 100.0%; | Pred. No. 1.9e-76; | | | | | |
| XX | Matches | 216: Conservative | 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
| XX | 41 | EYKEQIQIRVAVASGMQAAVELLSTLEKGVWHLGWTREVEALRLRTGSLAAKYNHPELT | 100 | | | | | | |
| XX | DB | 91 | EYKEQIQIRVAVASGMQAAVELLSTLEKGVWHLGWTREVEALRLRTGSLAAKYNHPELT | 150 | | | | | |
| XX | 01 | DDPSSFEAHQDEYIQLNLQPTLVDKRLVADYLDKCEEEELTIEDNNI | 160 | | | | | | |
| XX | DB | 151 | DLPSSFEAHQDEYIQLNLQPTLVDKRLVADYLDKCEEEELTIEDNNI | 210 | | | | | |

QY 161 ESGVRELLKRIYQKEMFSAFLNVLROTGNNELVQELTSGDCSESNAEITNLSQVDPGV 220
 |||||||
 DB 211 ESGVRELLKRIYQKEMFSAFLNVLROTGNNELVQELTSGDCSESNAEITNLSQVDPGV 270
 |||||||
 QY 221 EBOILLSTVQPNLEKEVGMENNSSESPADSSVVS 256
 |||||||
 DB 271 EBOILLSTVQPNLEKEVGMENNSSESPADSSVVS 306
 |||||||

RESULT 11
 ID AAU00297 standard; Protein: 267 AA.
 AC AAU00297;
 XX
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN5.
 XX
 KM Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KM graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KM haematologic disease; chronic neutropenia; myocardial infarction;
 KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5.
 XX
 OS Homo sapiens.
 XX
 PN WO200118208-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000WO-US24704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 XX
 PA (CURA-) CURAGEN CORP.
 PA (BIO) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR N-PSDB: AAS01150.
 XX
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 1; Page 33-35; 134pp; English.
 XX

The sequence represents interferon induced polypeptide, IFN5. IFN
 nucleic acids and polypeptides are useful for treating or preventing a
 pathology associated with IFN polypeptide in a human. They are useful for
 determining the presence of or predisposition to a disease associated
 with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 acids, polypeptides and antibodies are useful for diagnosis, prevention
 or treatment of variety of immunological and cell proliferative
 disorders, such as autoimmune diseases e.g. lupus erythematosus.
 immunodeficiency diseases such as acquired immunodeficiency syndrome
 (AIDS), graft rejection, viral infections including hepatitis and human
 immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 haematologic diseases such as aplastic anaemia and chronic neutropenia
 and cancer. In addition they are also useful for treating or preventing
 various disorders associated with cell death, including myocardial
 infarction, stroke, neurological diseases including Alzheimer's and
 Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 atrophy. IFN nucleic acids and polypeptides are also useful for
 identifying interferon-like proteins and interferon agonists, for

CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 SO Sequence 267 AA;
 Query Match 16.2%; Score 861; DB 22; Length 267;
 Best Local Similarity 96.6%; Pred. No. 1.6e-58;
 Matches 170; Conservative 2; Mismatches 4; Indels 0 Gaps 0;

QY 41 EYKEIOIQTAVATSGNMQAVELLSTLEKGVHMLGTRFEVALRRTGSPLAARYNNELT 100
 |||||||
 DB 87 EYKEIOIQTAVATSGNMQAVELLSTLEKGVHMLGTRFEVALRRTGSPLAARYNNELT 146
 |||||||
 QY 101 DLPSPFENAHDEYIQLNLLOPTLYDKLLRVDDYDKMBEELLTIEDRRNR1AAANN3N 160
 |||||||
 DB 147 DLPSPFENAHDEYIQLNLLOPTLYDKLLRVDDYDKMBEELLTIEDRRNR1AAANN3N 206
 |||||||

QY 161 ESGVRELLKRIYQKEMFSAFLNVLROTGNNELVQELTSGDCSESNAEITNLSQVDPGV 216
 |||||||
 DB 207 ESGVRELLKRIYQKEMFSAFLNVLROTGNNELVQELTSGDCSESNAEITNLSQVDPGV 262
 |||||||

RESULT 12
 ID AAU00299 standard; Protein: 271 AA.
 AC AAU00299;
 XX
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN7.
 XX
 KM Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KM graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KM haematologic disease; chronic neutropenia; myocardial infarction;
 KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN7.
 XX
 OS Homo sapiens.
 XX
 PN WO200118208-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000WO-US24704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 XX
 PA (CURA-) CURAGEN CORP.
 PA (BIO) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR N-PSDB: AAS01152.
 XX
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 1; Page 44-46; 134pp; English.
 XX

The sequence represents interferon induced polypeptide, IFN7. IFN
 nucleic acids and polypeptides are useful for treating or preventing a
 pathology associated with IFN polypeptide in a human. They are useful for

CC determining levels of IFN polypeptide or predisposition to a disease associated
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC acids, polypeptides and antibodies are useful for diagnosis, prevention
CC or treatment of variety of immunological and cell proliferative
CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC immunodeficiency diseases such as acquired immunodeficiency syndrome
CC (AIDS), graft rejection, viral infections including hepatitis and human
CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC hematologic diseases such as aplastic anaemia and chronic neuropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon antagonists, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans.

XX Sequence 271 AA:

Query Match 16.2%; Score 861; DB 22; Length 271;
Best Local Similarity 96.6%; Pred. No. 1,76,58;
Matches 170; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 41 EYKEQIORTVATSGNMQAVELLSTLEKGMHGTREVEAI RRGSPH AARYNPBLT 100
DB 91 EYKEQIORTVATSGNMQAVELLSTLEKGMHGTREVEAI RRGSPH AARYNPBLT 150
QY 101 DLPSPFENAHDEYLOLLNLDTLVKLLVRYDVKCHEEELTTEDKNA AARNNN 160
DB 151 DLPSPFENAHDEYLOLLNLDTLVKLLVRYDVKCHEEELTTEDKNA AARNNN 210
QY 161 ESGVRELLKRIYCKENMFSAFLNVLKOTGNNELVCELTGSDSESNAEITNLSQVD 216
DB 211 ESGVRELLKRIYCKENMFSAFLNVLKOTGNNELVCELTGSDSESNAEITNLSQVD 266

RESULT 13
AAU23654
ID AAU23654 standard; Protein; 448 AA.
XX
AC AAU23654;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #740.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227709.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-465566/50.
 DR N-PSDB: AAS41524.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 11: SEQ ID NO 1650; 1180pp; English.
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,

CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC Immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAD22915-AA023814 represent the novel human enzyme po-ypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 448 AA;
 Query Match 15.2%; Score 808; DB 22; Length 418;
 Best local similarity 42.4%; Pred. No. 4,9e-54;
 Matches 172; Conservative 76; Mismatches 126; Indels 5; Gaps 5;
 QY 480 PLPQILGLTASPGVGATKQAKAEERHLKCANIDAFITKENTDQKNGICQCKKF 539
 DB 11 PLPOVIGLTASPGTGASKLIDGAIINHVLQCANLDTWCMSPONCCPOLCHESQVCKQY 70
 QY 540 AIAADATREDPPEKLEIMTRIQTYCQSPMS-DGPTPTYQMAIQMKKAKKQKKER 598
 DB 71 NLCHRRSDDPEGDLTKLMDQIHDLHLEPBLSRKGTQMYEDQYVXLEAALADQOR 130
 QY 599 VCAEHLARKYNEALQINDTRIMIDAYTHLETFYNEEKDKFVIEDDSDEGDEYCKDE 658
 DB 131 VYALHLRRYNALLIHDIVRAVDALAALQDFYHRHRYKTKQL-----C----- 174
 QY 659 DEDDLKPKLDEIDRFLMTLFFENNKKMLKRLAENPEYNEKTKRLNTIMEQYTKIES 718
 DB 175 -----AERRILALFDQKKNELAHATGP-ENPKLEMLKILQOFQSSNS 219
 QY 719 ARGILFTKROSAYVALSOMITENEFKFAVGVYAHHLTAGSSSEKPK-CNHQKSVISKF 778
 DB 220 PGGILFTKROSASHLLMLQOQGLQTVDIRAQLLIAGNSOSTHM-CRQUCVIOKFE 279
 QY 779 RTGKINLIATVAEGGLDKEKIVIRGYLTNEIAMVQAGRADESTYIVIVANSGS 838
 DB 280 QDGTINLIVATVAEEGLDIPHCNVVYRGILTINISVQARGKAAQSVYAVATGGS 339
 QY 839 GVIEHETVNDREKMMYKRAHCVQNMKPEYAKHLLLEQOMSIEMK 884
 DB 340 RELKRELINLELTLMEOVAVAVKMDQAEYQAKIRDLQAAALTR 385
 RESULT 14
 ABB42219
 ID ABB42219 standard; Peptide; 166 AA.
 XX
 AC ABB42219;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #9725 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver.
 PS
 CC Claim 27; SEQ ID NO 34854; 639pp + sequence listing; English.
 CC
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 CC
 SQ Sequence 166 AA;
 Query Match 14.7%: Score 783; DB 22; Length 166;
 Best Local Similarity 93.3%: Pred. No. 9.1e-53;
 Matches 152; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MSNGSTDENFRYLISCFARAKMYIOVEPVLDYTLFLPAEYKEQIORTVATSGNMAVE 60
 DB 1 MSNGSTDENFRYLISCFARAKMYIOVEPVLDYTLFLPAEYKEQIORTVATSGNMAVE 60
 QY 61 LLSTLEKGVNHLGWTREVEALRRTGSPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
 DB 61 LLSTLEKGVNHLGWTREVEALRRTGSPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
 QY 121 LQPTLVDKLLVRDYLDKCMEELLTIEDNRNIAAENNGNSG 163
 DB 121 LQPTLVDKLLVRDYLDKCMEELLTIEDNRNIVGVCSDGASLLG 163
 RESULT 15
 AAM63104
 ID AAM63104 standard; Protein: 166 AA.
 XX
 AC AAM63104;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35209.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN W0200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 XX brains.
 PS
 CC Example 4; SEQ ID NO: 35209; 650pp + Sequence Listing; English.
 CC
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 CC
 SQ Sequence 166 AA;
 Query Match 14.7%: Score 783; DB 22; Length 166;
 Best Local Similarity 93.3%: Pred. No. 9.1e-53;
 Matches 152; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MSNGSTDENFRYLISCFARAKMYIOVEPVLDYTLFLPAEYKEQIORTVATSGNMAVE 60
 DB 1 MSNGSTDENFRYLISCFARAKMYIOVEPVLDYTLFLPAEYKEQIORTVATSGNMAVE 60
 QY 61 LLSTLEKGVNHLGWTREVEALRRTGSPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
 DB 61 LLSTLEKGVNHLGWTREVEALRRTGSPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
 QY 121 LQPTLVDKLLVRDYLDKCMEELLTIEDNRNIAAENNGNSG 163
 DB 121 LQPTLVDKLLVRDYLDKCMEELLTIEDNRNIVGVCSDGASLLG 163

Search completed: May 8, 2003, 15:52:39
 Job time : 86 secs

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:578526"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: PCMV-SPOrt6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."
 BASE COUNT 383 a 161 c 229 g 240 t
 ORIGIN

Query Match 27.5%; Score 926.4; DB 14; Length 1013;
 Best Local Similarity 98.3%; Pred. No. 4,4e+188;
 Matches 989; Conservative 0; Mismatches 11; Indels 6; Gaps 5;

1752 GAAAAACCAATATGACGACCATGCAAGAGTTGGCCATGAGATGCAAGAGAGAGA 1811
 |||||||
 Db 1 GAAAAACCAATATGACGACCATGCAAGAGTTGGCCATGAGATGCAAGAGAGA 60
 |||||||

1812 TCCATTAAAGAACTCTAGAAATATGACAAAGATCAAACTTATCTCAATGAG 1871
 |||||||
 Db 61 TCCATTAAAGAACTCTAGAAATATGACAAAGATCAAACTTATCTCAATGAG 120
 |||||||

1872 TCCATGTAGATTTTGAACCTCAACCTATGACAAATGSCCATTCAAATGAAAAA 1931
 |||||||
 Db 121 TCCATGTAGATTTTGAACCTCAACCTATGACAAATGSCCATTCAAATGAAAAA 180
 |||||||

1932 AGCTGCAAAAAAGGAAATCGCAAGAGCTTTGTGCAAGCATTTGAGGAGATCA 1991
 |||||||
 Db 181 AGCTGCAAAAAAGGAAATCGCAAGAGCTTTGTGCAAGCATTTGAGGAGATCA 240
 |||||||

1992 TGAGCCCTACAAATTAATGACAACTTGAATGATAGAATGCTTACTATCTGAAAC 2051
 |||||||
 Db 241 TGAGCCCTACAAATTAATGACAACTTGAATGATAGAATGCTTACTATCTGAAAC 300
 |||||||

2052 TTTCTATATGAAGAAAGATAGAAGTTTGGAGTCAAGAAAGATGATGATGAGG 2111
 |||||||
 Db 301 TTTCTATATGAAGAAAGATAGAAGTTTGGAGTCAAGAAAGATGATGATGAGG 360
 |||||||

2112 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2171
 |||||||
 Db 361 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 |||||||

2172 ACTGATGAAACAGATAGATTTCTCATGACTTATTTTGAACCAATTAATGATGAA 2231
 |||||||
 Db 421 ACTGATGAAACAGATAGATTTCTCATGACTTATTTTGAACCAATTAATGATGAA 480
 |||||||

2232 AAGGTGGCTGAAGAACCCAGATATGAAAAAGCCACCAATTAATGAAATAT 2291
 |||||||
 Db 481 AAGGTGGCTGAAGAACCCAGATATGAAAAAGCCACCAATTAATGAAATAT 540
 |||||||

2292 AATGAGCAATATAGTACGATGAGATGACAGAGGATTAATTTTACAAAAACAG 2351
 |||||||
 Db 541 AATGAGCAATATAGTACGATGAGATGACAGAGGATTAATTTTACAAAAACAG 600
 |||||||

2352 ACAGAGTGCATATGCGCTTCCAGTGAATGATGAAAAAGAAAAATTTCTGAAGTAG 2411
 |||||||
 Db 601 ACAGAGTGCATATGCGCTTCCAGTGAATGATGAAAAATTTCTGAAGTAG 660
 |||||||

2412 AGTCAAAAGCCCACTATGATGAGCTGGACACAGAGTGAAGCCATGACACA 2471
 |||||||
 Db 661 AGTCAAAAGCCCACTATGATGAGCTGGACACAGAGTGAAGCCATGACACA 720
 |||||||

2472 GAATGAACAAAAAGAGTATTAGTAATTTGACATGAAAAATGCAATCTGATAGG 2531
 |||||||
 Db 721 GAATGAACAAAAAGAGTATTAGTAATTTGACATGAAAAATGCAATCTGATAGG 780
 |||||||

2532 TACACAGTGGCAGAGAGAGTCTGATATTAAGAAATCAACATTTGATCCGTTATGG 2591
 |||||||
 Db 781 TACACAGTGGCAGAGAGAGTCTGATATTAAGAAATCAACATTTGATCCGTTATGG 840
 |||||||

2592 TCTCGTCAACCAATGAAATAGCATGTGTCAGAGCCCGCTGTCAGAGCCAGACTGATGAG 2651
 |||||||

|||||
 Db 841 TCTGCTCA-CATGCAAAATAGCATGCTCAGAGCCCGTGTGACCCAGAGCTCATGAGAG 899
 |||||||

2652 CACCTACCTCTCGTTGCTCAGAGTGTTCAGAGATTCGACAGATCAATATGA 2711
 |||||||
 Db 900 CACCTACCTCTCGTTGCTCAGAGTGTTCAGAGATTCGACAGATCAATATGA 956
 |||||||

2712 TTTCCGAGAGAGATGATGATTAAGCTATACATTTGTTGTTCAAAAT 2757
 |||||||
 Db 957 TTTCCGAGAGAGATGATGATTAAGCTATACATTTGTTGTTCAAAAT 1000
 |||||||

RESULT 2

BM467983 1115 bp mRNA linear EST 05-FEB-2002
 LOCUS BM467983

DEFINITION AGENCOURT 6437921 NIH_MGC_71 Homo sapiens cdna clone IMAGE:5532884
 5', mRNA sequence.

ACCESSION BM467983
 VERSION BM467983.1 GI:18517025

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1115)
 NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@email.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>
 Plate: L14M1216 Row: n Column: 21

High quality sequence stop: 690.

FEATURES

source

1..1115

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5532884"
 /clone_lib="NIH_MGC_71"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: PCMV-SPOrt6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

BASE COUNT 326 a 221 c 218 g 277 t 73 others

ORIGIN

Query Match 24.0%; Score 806.8; DB 13; Length 1115;
 Best Local Similarity 99.6%; Pred. No. 2e+162;
 Matches 808; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2555 TGGATATTAAAGAAATGTAACATTTGATTCGTTATGCTGCTGACCAATGAATAGCA 2614
 |||||||
 Db 2 TGGATATTAAAGAAATGTAACATTTGATTCGTTATGCTGCTGACCAATGAATAGCA 61
 |||||||

2615 TGGTCCAGCCCGTGGTGGAGCCAGAGTGTATGAGAGCAGCTAGCTGTTGCTACA 2674
 |||||||
 Db 62 TGGTCCAGCCCGTGGTGGAGCCAGAGTGTATGAGAGCAGCTAGCTGTTGCTACA 121
 |||||||

2675 GTGGTTCAGAGTATTCGACATGAGACAGTAAATGATTTCCGAGAGAGATGATATA 2734
 |||||||
 Db 122 GTGGTTCAGAGTATTCGACATGAGACAGTAAATGATTTCCGAGAGAGATGATATA 181
 |||||||

2735 AAGCATATCATTTGTTCAAAATATGAACAGAGAGATGATGATGATTTTGAAT 2794
 |||||||
 Db 182 AAGCATATCATTTGTTCAAAATATGAACAGAGAGATGATGATGATTTTGAAT 241
 |||||||

```

OY 2795 TACAGATGCAAGATTAATGAGAAAAAGAAATGAAACGAGAAATATTCACAGACTT 2854
    |||||||
DB 242 TACAGATGCAAGATTAATGAGAAAAAGAAATGAAACGAGAAATATTCACAGACTT 301
OY 2855 ACAAGATTAACCCATCACTAATAACTTCTTGGAAAAAGAGAGTGTCTTAGCTCTT 2914
    |||||||
DB 302 ACAAGATTAACCCATCACTAATAACTTCTTGGAAAAAGAGAGTGTCTTAGCTCTT 361
OY 2915 CTGGGAGAGATATTCATTAATGAGAAAAATGCATCAGCTCATATGATGATGAGAAATTC 2974
    |||||||
DB 362 CTGGGAGAGATATTCATTAATGAGAAAAATGCATCAGCTCATATGATGATGAGAAATTC 421
OY 2975 AGGAACCTTACATTTGTAAGAGAAAAACAAAGCAGCAAGAGTGTCTTAGCTCTTCAA 3034
    |||||||
DB 422 AGGAACCTTACATTTGTAAGAGAAAAACAAAGCAGCAAGAGTGTCTTAGCTCTTCAA 481
OY 3035 TAAATGAGAAATCATCTGCAAAATGTGGCAGGCTTGGAGAAATGATGATGAGCAAG 3094
    |||||||
DB 482 TAAATGAGAAATCATCTGCAAAATGTGGCAGGCTTGGAGAAATGATGATGAGCAAG 541
OY 3095 GCTTAGATTTGCTTGTCTGCAAAATGAGAAATTTTGTGAGTGTCTTCAATATTCAA 3154
    |||||||
DB 542 GCTTAGATTTGCTTGTCTGCAAAATGAGAAATTTTGTGAGTGTCTTCAATATTCAA 601
OY 3155 CAAAGAAACATACAAAGAGTGGTAGAATTAACCTATCAGATTTCCCAATCTTGACTAT 3214
    |||||||
DB 602 CAAAGAAACATACAAAGAGTGGTAGAATTAACCTATCAGATTTCCCAATCTTGACTAT 661
OY 3215 CAGAAATGCTTTTATTTAGTATGAGAGATTAAGACTTATGATGATGATGATGATGAT 3274
    |||||||
DB 662 CAGAAATGCTTTTATTTAGTATGAGAGATTAAGACTTATGATGATGATGATGATGAT 721
OY 3275 TATCAGTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3334
    |||||||
DB 722 TATCAGTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781
OY 3335 TAAGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3365
    |||||||
DB 782 TAAGATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812

```

```

RESULT 3
BO960157 870 bp mRNA line EST 21-AUG-2002
LOCUS BO960157
DEFINITION AGENCOURT_8949486 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6472264
VERSION BO960157.1 GI:22375635
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE NIH-MGC http://mgi.nhl.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apds-remail.nih.gov

```

```

FEATURES
source
1..870
location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:6472264"

```

```

/c1one_1lb="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Origin: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average Insert size 2.1 kb."
BASE COUNT 319 a 131 c 204 g 214 t 2 others
ORIGIN

```

```

Query Match 23.8% Score 801.8; DB 14; Length 870;
Best Local Similarity 98.4%; Pred. No. 2.3e-161;
Matches 851; Conservative 0; Mismatches 9; Indels 5; Gaps 4;

```

```

OY 1940 AAAAAAGAAATGCAAGAGTGTGTCAGCAAACTTTGAGAGATGATGAGAGCC 1999
    |||||||
DB 1 AAGAGAGAAATGCAAGAGTGTGTCAGCAAACTTTGAGAGATGATGAGAGCC 59
OY 2000 TACAATTAATGACAAATTCGAATGATGATGATGATGATGATGATGATGATGATGAT 2059
    |||||||
DB 60 TACAATTAATGACAAATTCGAATGATGATGATGATGATGATGATGATGATGATGAT 119
OY 2060 ATGAAGAGAAATGATGAGAGTGTGTCAGCAAACTTTGAGAGATGATGAGAGCC 2119
    |||||||
DB 120 ATGAAGAGAAATGATGAGAGTGTGTCAGCAAACTTTGAGAGATGATGAGAGCC 179
OY 2120 ATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2179
    |||||||
DB 180 ATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239
OY 2180 AAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2239
    |||||||
DB 240 AAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 229
OY 2240 CTGAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2299
    |||||||
DB 300 CTGAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359
OY 2300 AATATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359
    |||||||
DB 360 AATATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
OY 2360 CATATGCGCTTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2419
    |||||||
DB 420 CATATGCGCTTCCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
OY 2420 CCCACCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2479
    |||||||
DB 480 CCCACCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
OY 2480 AAAAAAGATCTTAAATTTGCACTGCAAAATCAATCTGCTATTCCTATCCAG 2539
    |||||||
DB 540 AAAAAAGATCTTAAATTTGCACTGCAAAATCAATCTGCTATTCCTATCCAG 599
OY 2540 TGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2599
    |||||||
DB 600 TGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
OY 2600 CCAATGAATATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2659
    |||||||
DB 660 CCAATGAATATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719
OY 2660 TCGTGTGCTGACAGT-GGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 2718
    |||||||
DB 720 TCGTGTGCTGACAGTGTGTCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 779
OY 2719 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2775
    |||||||
DB 780 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
    |||||||
OY 2776 GCTCATAGATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2800
    |||||||
DB 840 GCTCATAGATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864

```

RESULT 4
 LOCUS B0772836/c 755 bp mRNA linear EST 26-JUL-2002
 DEFINITION UI-H-FEO-bdn-c-04-0-UI-s1 NCI-CGAP_FEO Homo sapiens cDNA clone
 ACCESSION B0772836
 VERSION B0772836.1 GI:21981312
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 755)
 NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: James Martin
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@iowa.edu;
 Seq primer: M13 FORWARD
 POLYA-Yes.
 FEATURES
 source Location/Qualifiers
 1..755
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-H-FEO-bdn-c-04-0-UI-s1"
 /clone_11b="NC1-CGAP_FEO"
 /tissue_type="Chondrosarcoma Cell Line"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73-Pac (Pharmacia) with a modified
 polylinker. Site 1: EcoR I; Site 2: Not I; NC1-CGAP_FEO is
 a cDNA library containing the following tissue(s): a pool
 of 3 chondrosarcoma cell lines (grade 2) The library was
 constructed according to Boudelo, Lemm and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I and cloned directionally
 into pT73-Pac vector. The (unlabeled) ide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 CGCTACGGAC. The cell lines was provided by Dr James Martin
 of University of Iowa.
 TAG_L1b-UI-H-FEO
 TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
 TAG_SEQ=CGCTACGGAC"
 BASE COUNT 207 a 141 c 115 g 290 t 2 oth:is
 ORIGIN
 Query Match 21.4%; Score 718.8; DB 14; Length 755;
 Best Local Similarity 99.3%; Pred. No. 1.4e-143;
 Matches 731; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 QY 2630 GTGAGCCAGAGCTGATGAGACACCTAGCTCTGTCGACAGTGTTCAGGAGTTA 2689
 DB 754 GTGAGCCAGAGCTGATGAGACACCTAGCTCTGTCGACAGTGTTCAGGAGTTA 696
 QY 2690 TCGACATGAGACAGTATGATTCGCGAGAGATGATGATTAAGTATACATTG 2749
 DB 695 TCGACATGAGACAGTATGATTCGCGAGAGATGATGATTAAGTATACATTG 636
 QY 2750 TTCAAAATATGAACAGAGAGATGATGATTAAGTATGATTAACATGCAAAAGTA 2809
 DB 635 TTCAAAATATGAACAGAGAGATGATGATTAAGTATGATTAACATGCAAAAGTA 576

QY 2810 TAATGAAAAGAAAATGAAAACCAAGAGAAATATTCACAGATTACAAAGAT 2869
 DB 575 TAATGAAAAGAAAATGAAAACCAAGAGAAATATTCACAGATTACAAAGAT 516
 QY 2870 CACTATATACCTTCTCTTGCAAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2929
 DB 515 CACTATATACCTTCTCTTGCAAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 456
 QY 2930 ATGTAAATGAGAAAATGATACGTCATGATGACCCCGAATTCAGAACTTACATG 2989
 DB 455 ATGTAAATGAGAAAATGATACGTCATGATGACCCCGAATTCAGAACTTACATG 396
 QY 2990 TAAAGAAAACAAAGACACTGCAAAAAGAGTGGCGACTATCAAAATGAATGCTGAAATCA 3049
 DB 395 TAAAGAAAACAAAGACACTGCAAAAAGAGTGGCGACTATCAAAATGAATGCTGAAATCA 336
 QY 3050 TCTGCAATGTCGCCAGCTGTCGGGACAAATGATGCTGCAAAAAGCTTACATTCCTT 3109
 DB 335 TCTGCAATGTCGCCAGCTGTCGGGACAAATGATGCTGCAAAAAGCTTACATTCCTT 276
 QY 3110 GTCTCAAAATAGAAATTTTGTAGTGTTCACAAAATATTCACAAAAGAAATATACA 3169
 DB 275 GTCTCAAAATAGAAATTTTGTAGTGTTCACAAAATATTCACAAAAGAAATATACA 216
 QY 3170 AAAAGTGGGTAGAAATTTACCTATCAATTTCCCAATCTTGACTATTCAGAAATGCTGCTTAT 3229
 DB 215 AAAAGTGGGTAGAAATTTACCTATCAATTTCCCAATCTTGACTATTCAGAAATGCTGCTTAT 156
 QY 3230 TTAGTATGAGAGATGATGACCTGATGAGATTTCTTTAAATATCATCATTAACAT 3289
 DB 155 TTAGTATGAGAGATGATGACCTGATGAGATTTCTTTAAATATCATCATTAACAT 96
 QY 3290 TAAATGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3349
 DB 95 TAAATGATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 36
 QY 3350 TGATGTTTACTCTG 3365
 DB 35 TGATGTTTACTCTG 20
 RESULT 5
 BP337464
 LOCUS BP337464 1035 bp mRNA linear EST 22-NOV-2000
 DEFINITION 602035199F1 NC1-CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183126
 5', mRNA sequence.
 ACCESSION BP337464
 VERSION BP337464.1 GI:11283715
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1035)
 NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM9498 row: n column: 23
 High quality sequence start: 5
 High quality sequence stop: 695.
 FEATURES
 source Location/Qualifiers
 1..1035
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:4183126"
 /clone_lib="NCI CGAP Brn64"
 /tissue_type="gliblastoma with EGFR amplification"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pCMV-Sport6; Site: 1; Not:
 Site: 2; Salt: Cloned unidirectionally; Primer: Cligo dt.
 Average insert size 1.57 kb. Constructed by Life
 Technologies. Note: this is a NCI CGAP library."
 BASE COUNT 353 a 221 c 243 g 218 t
 ORIGIN

Query Match 21.1%; Score 711.6; DB 12; Length 1035;
 Best Local Similarity 97.2%; Pred. No. 5,3e-142;
 Matches 756; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 1049 AGAATGTGCGAGAGAGATCCCGGAGCCAGCACTCCAGTCCAGCCCTACCAATATGG 1108
 DB 4 AGAATGTGCGAGAGAGATCCCGGAGCCAGCACTCCAGTCCAGCCCTACCAATATGG 63
 QY 1109 AAGTTCGCCAGCGCTTGGAGGAGATATATCATCTGCGCTCCCTAGGAGATG 1168
 DB 64 AAGTTCGCCAGCGCTTGGAGGAGATATATCATCTGCGCTCCCTAGGAGATG 123
 QY 1169 GAAAAACGAGTGTCTTACATTCGCAAGATCCTTACAGCAAGAA'AAAAAGCAT 1228
 DB 124 GAAAAACGAGTGTCTTACATTCGCAAGATCCTTACAGCAAGAA'AAAAAGCAT 183
 QY 1229 CTGAGCTGGAAAGTTATAGTCTTGTCAATAGTACTGCTAGTTGA'AGCTCTCC 1288
 DB 184 CTGAGCTGGAAAGTTATAGTCTTGTCAATAGTACTGCTAGTTGA'AGCTCTCC 243
 QY 1289 GGAAGAGTTCACACATTTTGAACAATGATGCTGTTATGATTAAGTGGTGA 1348
 DB 244 GGAAGAGTTCACACATTTTGAACAATGATGCTGTTATGATTAAGTGGTGA 303
 QY 1349 CCCAATGAAATATATCTTCGAGAGTGTCAAGTCTGTGATATATATACAGTACG 1408
 DB 304 CCCAATGAAATATATCTTCGAGAGTGTCAAGTCTGTGATATATATACAGTACG 363
 QY 1409 CTCGAATCTTGAACCTCCCTCTTAACCTTGAACCAATGAGATGCTGTTCAAT 1468
 DB 364 CTCGAATCTTGAACCTCCCTCTTAACCTTGAACCAATGAGATGCTGTTCAAT 423
 QY 1469 TGTCAACTTTTCCCTCATATATGATGATGATGATGATGATGATGATGATGATG 1528
 DB 424 TGTCAACTTTTCCCTCATATATGATGATGATGATGATGATGATGATGATGATG 483
 QY 1529 ATATATATCATGATGAGCATATATATGATGATGATGATGATGATGATGATGATG 1588
 DB 484 ATATATATCATGATGAGCATATATATGATGATGATGATGATGATGATGATGATG 543
 QY 1589 AAAACAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1648
 DB 544 AAAACAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 603
 QY 1649 GAGGGGCGAG 1708
 DB 604 GAGGGGCGAG 662
 QY 1709 ATGATTTTATTAAGTGTGAAGAAAGCTTGTATGATGATGATGATGATGATGATG 1768
 DB 663 ATGATTTTATTAAGTGTGAAGAAAGCTTGTATGATGATGATGATGATGATGATG 722
 QY 1769 AGCGATGCAAGAGTTGCGATGATGATGATGATGATGATGATGATGATGATGATG 1826
 DB 723 GCGCATGCGAGAG- TTGGGGGCTGCGATGCGAGC- GAGAGAGATCCCTT'AGGGAAA 778
 RESULT 6 BF983236 729 bp mRNA line-1 EST 23-JAN-2001
 LOCUS BF983236
 DEFINITION 602305873F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4397083 5',

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC <http://mgs.nci.nih.gov/>
 1 (bases 1 to 729)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

CDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LNL0097 Row: a Column: 20
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4397083"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-Sport6;
 Site: 1; Not: Site: 2; Salt: Cloned unidirectionally;
 oligo-primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 240 a 137 c 184 g 168 t
 ORIGIN

Query Match 20.4%; Score 687; DB 12; Length 729;
 Best Local Similarity 98.9%; Pred. No. 9,2e-137;
 Matches 723; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 586 GAGGAGGAACCTGTGACATTAAGACAGAAACCGATGCTGCGAAGAAACATGA 645
 DB 1 GAGGAGGAACCTGTGACATTAAGACAGAAACCGATGCTGCGAAGAAACATGA 60
 QY 646 AATGAATCAGTGTGAAGAGCTACTAAAAAGATTTGCGAAGAAAGAACTGG:TCCT 705
 DB 61 AATGAATCAGTGTGAAGAGCTACTAAAAAGATTTGCGAAGAAAGAACTGG:TCCT 120
 QY 706 GCATTTTCAATGTTCTTCTGTCAAACAGGAACAATGAATGTTCCAGAGATTACAGGC 765
 DB 121 GCATTTTCAATGTTCTTCTGTCAAACAGGAACAATGAATGTTCCAGAGATTACAGGC 180
 QY 766 TGTGATGCTGCAAGAAAGATGAGATGGAATTTATCAAGATGATGATG:CTCA 825
 DB 181 TGTGATGCTGCAAGAAAGATGAGATGGAATTTATCAAGATGATGATG:CTCA 240
 QY 826 GTGGAAGACAACTCTTTCACACAGTTCAGCCAAATCTGAGAGAGAGCTG:GGGGC 885
 DB 241 GTGGAAGACAACTCTTTCACACAGTTCAGCCAAATCTGAGAGAGAGCTG:GGGGC 300
 QY 886 ATGAGAAATTAATCATCATCATCTTTTGCAGATTTCTTGATGTTTCAAG:CAAGC 945
 DB 301 ATGAGAAATTAATCATCATCATCTTTTGCAGATTTCTTGATGTTTCAAG:CAAGC 360
 QY 946 ACAAGTTTGGCAAGGAAGTGTCAAGCTGTAGATGAAAGCTTGGACATPAACAGCAAC 1005
 DB 361 ACAAGTTTGGCAAGGAAGTGTCAAGCTGTAGATGAAAGCTTGGACATPAACAGCAAC 420
 QY 1006 ATGGGAGTATTCAGGCACCATGGGAAGTATTCAGATGAAGAGATGTGCA:CAAGA 1065
 |||||||

DB 421 ATGGGAGATGATTCAGGACGACGATGGAAGTATTCAGATGAGAGA:AAATG133CAGACAGA 480

QY 1066 GCATCCCGGAGCCAGCAACTCCAGCTCAGGCGCTTACCAATGAGACTT3:CVAGCCACCC 1125

DB 481 GCATCCCGGAGCCAGCAACTCCAGCTCAGGCGCTTACCAATGAGACTT3:CVAGCCACCC 539

QY 1126 TTGGAAGGAGAAATATCATCATCTGCTCCCTCAGGAGAGTGGAAAAA:AGAGTGGCT 1185

DB 540 TTGGAAGGAGAAATATCATCATCTGCTCCCTCAGGAGAGTGGAAAAA:AGAGTGGCT 598

QY 1186 GTTACATTCGCCAGGATCAGTACAGCAAGAAAAAGCATCGAGG:CGAAAGCTT 1245

DB 599 GTTACATTCGCCAGGATCAGTACAGCAAGAAAAAGCATCGAGG:CGAAAGCTT 657

QY 1246 ATAGTCTGTGTCATTAAGTACTGCTAGTGAACAGCTCTCCGCAAG:CGTTCCACCA 1305

DB 658 ATAGTCTGTGTCATTAAGTACTGCTAGTGAACAGCTCTCCGCAAG:CGTTCCACCA 717

QY 1306 TTTTGAAGAA 1316

DB 718 TTTTGAAGAA 728

RESULT 7
BG741146 827 bp mRNA linear EST 15-MAY-2001
LOCUS 602631817F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776869 5'

DEFINITION mRNA sequence.

ACCESSION BG741146

VERSION BG741146.1 GI:14051799

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 827)
NIR-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov

TITLE Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

COMMENT Plate: LAM10630 row: b column: 06
High quality sequence stop: 763.

FEATURES
Location/Qualifiers
1..827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4776869"
/clone_id="NCI_CGAP_Skn3"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: skin; Vector: pCMV-SK3; Site: NotI;
Site-2: SalI; Cloned unidirectionally; primer: oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 254 a 186 c 204 g 183 t

ORIGIN

Query Match 19.4% Score 652.8; DB 12; Length 827;
Best Local Similarity 96.1% Pred. NO. 2.1e-129;
Matches 798; Conservative 0; Mismatches 17; Indels 15; Gaps 12;

QY 426 CGGACGCTCTGGCCGCGCCCTACATGAGACCTGAGCTCAGGACTT3:CTCTCATC 485

DB 1 CGGACGCTCTGGCCGCGCCCTACATGAGACCTGAGCTCAGGACTT3:CTCTCATC 59

QY 486 GTTGAAGAGGATGATGATATCTCCAACTGCGAAGCTCTTCAG:TCGACTCTGG 544

DB 60 GTTGAAGAGGATGATGATATCTCCAACTGCGAAGCTCTTCAG:TCGACTCTGG 119

QY 545 TGGACAGCTTCTAGTTGAGACGCTCTTGAGTAATGATGAGAGGAACTG1:GAGCA 604

DB 120 TGGACAGCTTCTAGTTGAGACGCTCTTGAGTAATGATGAGAGGAACTG1:GAGCA 179

QY 605 TTGAAGACAGAAACCGGATGCTGCTGAGAAACCAATGAAATGAAATCAGTCT:TAAGAG 664

DB 180 TTGAAGACAGAAACCGGATGCTGCTGAGAAACCAATGAAATGAAATCAGTCT:TAAGAG 239

QY 665 AGCTACTAAAAAGATTTGTGCAAGAAAGAACTGGTTCTTGATTTTGTGATGTTCTC 724

DB 240 AGCTACTAAAAAGATTTGTGCAAGAAAGAACTGGTTCTTGATTTTGTGATGTTCTC 299

QY 725 GTCAACAGGAAACATGAACTTGTGCAAGAGTTAAAGGCTCTGATTTG--CT:GGAAG 782

DB 300 GTCAACAGGAAACATGAACTTGTGCAAGAGTTAAAGGCTCTGATTTGCTGCAAAAAG 359

QY 783 CAATGAGAGATTTAGAAATTTATCAAGATTGATGCTCTCAAGTGAAGAC:AACTTCT 842

DB 360 CAATGAGAGATTTAGAAATTTATCAAGATTGATGCTCTCAAGTGAAGAC:AACTTCT 419

QY 843 TTCACACAGATTCAGCCAAATCTGGA--GAAGAGAGCTGTGGGATGAGAAATACATCAT 901

DB 420 TTCACACAGATTCAGCCAAATCTGAGCAAGAGAGCTGTGGGATGAGAAATACATCAT 479

QY 902 CAGAAATCATCTTTTGA--GATTTCTCTGATTTTCAAGATGAGACACAAAGTTTGA--G 958

DB 480 CAGAAATCATCTTTTGA--GATTTCTCTGATTTTCAAGATGAGACACAAAGTTTGA--G 539

QY 959 AAGGAAGTGTACAGCTGCTTAGATGAAGAGCTTTGA--CATTAACAGACATGGCAGATGAT 1017

DB 540 AAGGAAGTGTACAGCTGCTTAGATGAAGAGCTTTGA--CATTAACAGACATGGCAGATGAT 599

QY 1018 TCAGGACACATGAGGAAAGTATTCAGATGA--GAGAAATGGCAGCAAGATC:CCGGA 1076

DB 600 TCAGGACACATGAGGAAAGTATTCAGATGA--GAGAAATGGCAGCAAGATC:CCGGA 659

QY 1077 GCCAAGATCCAGCTCAGGCGCTTACCAAAATG--AAATGGCCAGCCAGCTTGAAGG 1134

DB 660 GCCAAGATCCAGCTCAGGCGCTTACCAAAATG--AAATGGCCAGCCAGCTTGAAGG 718

QY 1135 AAGAAATATCATCTGCTGCTTACAGAGGAGTGGAAAAACCAAGTGGCTTTACATT 1194

DB 719 AAGAAATATCATCTGCTGCTTACAGAGGAGTGGAAAAACCAAGTGGCTTTACATT 777

QY 1195 -GCCAAGATCCAGCTCAGGCGCTTACCAAAATG--AAATGGCCAGCCAGCTTGAAGG 124

DB 778 AGCCAGGATCAGTACAGCAAGAAACCAAGCATCTGAGCTTGAAGAG 827

RESULT 8
BF686405 781 bp mRNA linear EST 22-DEC-2000
LOCUS 602144786F1 NIR_MGC_46 Homo sapiens cDNA clone IMAGE:4304805 5'

DEFINITION mRNA sequence.

ACCESSION BF686405

VERSION BF686405.1 GI:11971813

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 781)
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIR-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov

TITLE Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1CML170 row: 1 column: 22
 High quality sequence stop: 731.

FEATURES

source

1..781
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4304805"
 /clone_lib="NIH_MGC_46"
 /tissue_type="Telomysarcoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pORF7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dt priming. 1. directionally cloned
 into EcoRI/XhoI sites using the following 4' 5' adaptor:
 GCCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

BASE COUNT 235 a 168 c 204 g 174 t

ORIGIN

Query Match 19.4%; Score 651.8; DB 12; Length 781;

Best Local Similarity 99.4%; Pred. No. 3.3e-129;

Matches 675; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

385 CTGTGGTGGAGCTGGGAATTCGTGGAGGCCCTCCGAGAACCGGACGCCTTGGCCGCC 444

5 CCTGTGGGAGCTGGGAATTCGTGGAGGCCCTCCGAGAACCGGACGCCTTGGCCGCC 64

445 CGGTATATGAACTGAGTCAAGCTCAAGGACTGCTCCATCGTTTGAAGAGTGTCTATGAT 504

65 CGGTATATGAACTGAGTCAAGGACTGCTCCATCGTTTGAAGAGTGTCTATGAT 124

505 GAATATCTCCAACTGCTGAACCTCTTCAAGCCACTCGTGTGAGCAAGCTTATGATGAG 564

125 GAATATCTCCAACTGCTGAACCTCTTCAAGCCACTCGTGTGAGCAAGCTTATGATGAG 184

565 GAGCTGTGGATTAAGTCAAGTGAAGGAACTGTTGACAAATGGAACAAACCGGAT 624

185 GAGCTG-TGGATTAAGTCAAGTGAAGGAACTGTTGACAAATGGAACAAACCGGAT 243

625 GCTGCTGCAAGAAACATGGAATGATCAAGTGTAAAGAGCTACTAAAGAGATGTG 684

244 GCTGCTGCAAGAAACATGGAATGATCAAGTGTAAAGAGCTACTAAAGAGATGTG 303

685 CAGAAAGAAACCTGTTCTGATTCGTAATGTTCTTGTCAACAGTAAACATGAA 744

304 CAGAAAGAAACCTGTTCTGATTCGTAATGTTCTTGTCAACAGTAAACATGAA 363

745 CTGTGCAAGAGTAAACAGGCTCTGATTCGTAAGAGCATGACAGATAGAAATTA 804

364 CTGTGCAAGAGTAAACAGGCTCTGATTCGTAAGAGCATGACAGATAGAAATTA 423

805 TCACAAGTGAATGCTCTCAAGTGAAGCAACTCTTCAACACACAGTAGCCAAAT 864

424 TCACAAGTGAATGCTCTCAAGTGAAGCAACTCTTCAACACACAGTAGCCAAAT 483

865 CTGGAGAGGAGTCTGGGGCATGAGAAATCACTCATGCAATCATCTTGGAGATCT 924

484 CTGGAGAGGAGTCTGGGGCATGAGAAATCACTCATGCAATCATCTTGGAGATCT 543

925 TGTGTAGTTTCAAGATCAGACAGATTTGGCAGAGGAAGTGTACCTTTAGATGAA 984

544 TGTGTAGTTTCAAGATCAGACAGATTTGGCAGAGGAAGTGTACCTTTAGATGAA 603

985 AGTCTTGGACATTAACAG-CACATGGGCACTGATTCAGGACCATGGGAATGATTCA 1043

604 AGTCTTGGACATTAACAGCAACATGGGCACTGATTCAGGACCATGGGAATGATTCA 663

1044 TGAAGAGATGTGGCAGCA 1062
 664 TGAAGAGATGTGGCAGCA 682

RESULT 9

LOCUS

BM476961

DEFINITION

AGENCOURT 6481569 NIH_MGC_71 Homo sapiens cDNA IMAGE:5555523

ACCESSION

BM476961

VERSION

BM476961.1 GI:18526003

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 1046)

AUTHORS

NIH-MGC <http://imgc.ncl.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1CML170 row: n column: 04

High quality sequence stop: 622.

Location/Qualifiers

1..1046

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5555523"

/clone_lib="NIH_MGC_71"

/tissue_type="Telomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb.

BASE COUNT 318 a 217 c 227 g 282 t 2 others

ORIGIN

Query Match 18.6%; Score 626; DB 13; Length 1046;

Best Local Similarity 97.7%; Pred. No. 1.2e-123;

Matches 635; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

172 TCGAATGGTATTCACAGACAGAAATTCCTATTCATCTGCTTCAGGACAGG 231

1 TCGAATGGTATTCACAGACAGAAATTCCTATTCATCTGCTTCAGGACAGG 60

232 GTGAAGAGTATACATCAGTGGAGGCTGTGCTGAGTACCTGCTGTGAGAG 291

61 GTGAAGAGTATACATCAGTGGAGGCTGTGCTGAGTACCTGCTGTGAGAG 120

292 GTGAAGAGTATCAGAGCAGTGGACCTCCGAGACATGACAGCAATTCAGTG 351

121 GTGAAGAGTATCAGAGCAGTGGACCTCCGAGACATGACAGCAATTCAGTG 180

352 CTGTGAGCAGCTTGGAGAGGAGTGTGGACCTTGGTGGACCTGGGAAATTCGTGAG 411

181 CTGTGAGCAGCTTGGAGAGGAGTGTGGACCTTGGTGGACCTGGGAAATTCGTGAG 240

412 GGCCTCCGAGAACCGGACCTCTGGGCGCGGCTGATCATGAACCTGAGTCAATGAG 471

241 GGCCTCCGAGAACCGGACCTCTGGGCGCGGCTGATCATGAACCTGAGTCAATGAG 300

472 TTGACCTTCATCGTTTGAAGAGCTCATGATATCTCAACTGCTTTCCTT 531

Db 301 TTGCGCTCTCCATGCTTTTGAGAACGCTCATGATGATATCTCCAACTGTTGAACCTCTT 360

QY 532 CAGCCCACTCTGTGGACAAAGCTTCTAGTAGAGAGCTTGGATAGTATATGAGAGAG 591

Db 361 CAGCCCACTCTGTGGACAAAGCTTCTAGTAGAGAGCTTGGATAGTATATGAGAGAG 420

QY 592 GAACCTGTGCAATTTGAGAGCAAGAACCGATGCTGCGCAAGAACCAATGGAATGAA 651

Db 421 GAACCTGTGCAATTTGAGAGCAAGAACCGATGCTGCGCAAGAACCAATGGAATGAA 480

QY 652 TCAGGTGTAGAGAGCTACTATAAAGAGATTGTGACAGAAAGAACTGCTTCTGCAATT 711

Db 481 TCAGGTGTAGAGAGCTACTATAAAGAGATTGTGACAGAAAGAACTGCTTCTGCAATT 540

QY 712 CTGAATGTTCTTGTCTCAACAGAGAAACATGAATGCTGCTCCAGAGATTAAAGGCTGAT 771

Db 541 CTGAATGTTCTTGTCTCAACAGAGAAACATGAATGCTGCTCCAGAGATTAAAGGCTGAT 600

QY 772 TGCTCAGAAAGCAATGACAGATTGAGAAATTTATCACAAGTTGATGCTG 821

Db 601 TGCTCAGAAAGCAATGACAGATTGATTTTATTTACTGAGAGAGATTCTT 950

RESULT 10
BE882040 1041 bp mRNA 11m EST 20-OCT-2000
LOCUS 601505326F2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906774 5',
DEFINITION mRNA sequence.
ACCESSION BE882040
VERSION BE882040.1 GI:10330816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 1041)
AUTHORS NIH-MGC http://mgs.scl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Inocyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9716 row: d column: 07
High quality sequence stop: 586.

FEATURES
Location/Qualifiers
1..1041

ORGANISM "Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3906774"
/clone_lib="NIH_MGC_71"
/tissue_type="telomysarcoma"
/lab_host="DH10B (phage-res-stand)"
/note="Organ: uterus; Vector: pCMV-Sport5; Site: 1; Note:
Site 2: Salt: Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 385 a 258 c 223 g 175 t
ORIGIN

Query Match 18.5% Score 622.8; DB 12; Length 1041;
Best Local Similarity 89.1%; Pred. No. 5.9e-123;
Matches 718; Conservative 0; Mismatches 82; Indels 6; Gaps 4;
QY 1923 GGAAGAAAAAGTGCAGAAAAAGAAATGCAAGAAAGCTGTTGTGCAATGATTGAG 1982
Db 1 GGAAGAAAAAGTGCAGAAAAAGAAATGCAAGAAAGCTGTTGTGCAATGATTGAG 60
QY 1983 GAAGTCAATGAGGCGCTTCAAAATTAA-TGACACAAATTGCAATGATAGATGATATCTC 2041

Db 61 GAAGTCAAGAGGCGCTTCAAAATTAAAGTACAGACAAATTTGAAATGATGATGATATCTC 120

QY 2042 ATCTTGAAGAACTTCTATATATGAGAGAAAGATTAAGAGTTTGCACTCATAGAGATGATA 2101

Db 121 ATCTTGAAGAACTTCTATATATGAGAGAAAGATTAAGAGTTTGCACTCATAGAGATGATA 180

QY 2102 GTGATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2161

Db 181 GTGATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 2162 AACCTTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2221

Db 241 AACCTTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 2222 AAATGTTGAAAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2281

Db 301 AAATGTTGAAAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360

QY 2282 GAAATACCATTAATGAGACATATCTAGAGCTGAGAGATCAGACAGAGAAATTAICTTTA 2341

Db 361 GAAATACCATTAATGAGACATATCTAGAGCTGAGAGATCAGACAGAGAAATTAICTTTA 420

QY 2342 CAAAAACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2401

Db 421 CAAAAACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 2402 CTGAAGTAGAGAGTCAAAAGCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 2461

Db 481 CTGAAGTAGAGATCAAAAGCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 540

QY 2462 CCATGACACAGATATACAAAAAGAAAGTATTAGTAAATTTGCACTGG-AAAAATCAAT 2520

Db 541 CCATGACACAGATATACAAAAAGAAAGTATTAGTAAATTTGCACTGGCAAAATAAT 600

QY 2521 CTGCTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580

Db 601 CTGCTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

QY 2581 ATCCGTTATGCTTCTGTCACCAATGAAATAGCCATGTCGACAGCCCGTGTGTCACACAGA 2640

Db 661 TCC---GCAGGCGCCAGACGACCAATGATAGGCAATAGGCGCAAGCCGCGCCCAAG-CAAC 717

QY 2641 CCTGATGAGAGCACTTACCTCTGCTG-CTCAGAGCTGTCAGAGTATGACATGATGA 2699

Db 718 TACTGAAAAAGACCAAAAGACGCGCTGCGCACAAAGGCTCACAGCAGCAACCAAGCGCA 777

QY 2700 GACAGTAAATGATTTCCGAGAGAGA 2725

Db 778 CACAGCGCAGACATCCAGAAACA 803

RESULT 11
AK018602 1197 bp mRNA linear HTC 19-JAN-2002
LOCUS AK018602
DEFINITION Mus musculus adult male cecum cDNA, RIKEN full-length enriched
library, clone:913009C22; homolog to MELANOMA DIFFERENTIATION
ASSOCIATED PROTEIN-5, full insert sequence.
ACCESSION AK018602
VERSION AK018602.1 GI:12858393
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male cecum cDNA to mRNA,
clone:913009C22.
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

| | | |
|------------|--------------------|---------------------------------------|
| | RESULT | 12 |
| BMA67774 | | |
| LOCUS | | |
| DEFINITION | BMA67774 | 1339 bp mRNA |
| | AGNCNCURT_643778.3 | 100% EST 05-FEB-2001 |
| | 5'... sequence. | Homo sapiens cDNA clone IMAGE:5532660 |

Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL12216 row: e column: 13
High quality sequence stop: 505.

| FEATURES | source | Location/Qualifiers |
|-------------|--------|--|
| | 1. | 1239 |
| | | /organism="Homo sapiens" |
| | | /db_xref="taxon:9606" |
| | | /clone IMAGE:5532660" |
| | | /clone_1id="NH_MGC_71" |
| | | /tissue_type="leiomyosarcoma" |
| | | /lab_host="DH10B (phage-resistant)" |
| | | /note="Organ: uterus; Vector: pCMV-Sp7b1"; Site_1: Nci1; |
| | | Site_2: Sal1; Cloned unidirectionally. Primer: Oligo dt. |
| | | Average insert size 2.1 kb. |
| BASE COUNT | 393 a | 285 c 243 g 318 t |
| ORIGIN | | |
| Query Match | 18.3% | Score 616.2; DB 13; Length 1239; |

| DB | Sequence | Position |
|-----|---|----------|
| Oy | AAACATGACCTGTCCAGAGATTAACAGAGCTGTGATTCCTCAGAAAGCAATG;A;TACATT | 795 |
| 736 | AAACATGACCTGTCCAGAGATTAACAGAGCTGTGATTCCTCAGAAAGCAATG;A;TACATT | 795 |
| 4 | AAACATGACCTGTCCAGAGATTAACAGAGCTGTGATTCCTCAGAAAGCAATG;A;TACATT | 63 |

| | | |
|----|--|-----|
| 0y | GAGAAATTATATACGAAGTTGATGTCCTCCTCAAGTCGAAAGCAACTTCTTTTCAACCTCCACAGCTT | 855 |
| 64 | GGAATTTATACGAAGTTGATGTCCTCCTCAAGTCGAAAGCAACTTCTTTTCAACCTCCACAGCTT | 123 |
| 0b | | |

856 CAGCCAAATCTGGAGAGAGAGTCTGGGGGCATGGAGAAATACCTCATCAGAAATCACTTTT 915
 124 CAGCCAAATCTGGAGAGAGAGTCTGGGGGCATGGAGAAATACCTCATCAGAAATCACTTTT 183

916 GCAGATTCCTGTGTTGACGAAATCAGACACAAAGTTGGCAGAGGAAGTGTCAAGTGC 975
 184 GCAGATTCCTGTGTTGACGAAATCAGACACAAAGTTGGCAGAGGAAGTGTCAAGTGC 243

976 TTAGTGAAGTCTTGACATACAGCAACATGGCGCATGATTGAGCACCATCTGGAGT 1035
 1040
 1045
 1050
 1055
 1060
 1065
 1070
 1075
 1080
 1085
 1090
 1095
 1100
 1105
 1110
 1115
 1120
 1125
 1130
 1135
 1140
 1145
 1150
 1155
 1160
 1165
 1170
 1175
 1180
 1185
 1190
 1195
 1200
 1205
 1210
 1215
 1220
 1225
 1230
 1235
 1240
 1245
 1250
 1255
 1260
 1265
 1270
 1275
 1280
 1285
 1290
 1295
 1300
 1305
 1310
 1315
 1320
 1325
 1330
 1335
 1340
 1345
 1350
 1355
 1360
 1365
 1370
 1375
 1380
 1385
 1390
 1395
 1400
 1405
 1410
 1415
 1420
 1425
 1430
 1435
 1440
 1445
 1450
 1455
 1460
 1465
 1470
 1475
 1480
 1485
 1490
 1495
 1500
 1505
 1510
 1515
 1520
 1525
 1530
 1535
 1540
 1545
 1550
 1555
 1560
 1565
 1570
 1575
 1580
 1585
 1590
 1595
 1600
 1605
 1610
 1615
 1620
 1625
 1630
 1635
 1640
 1645
 1650
 1655
 1660
 1665
 1670
 1675
 1680
 1685
 1690
 1695
 1700
 1705
 1710
 1715
 1720
 1725
 1730
 1735
 1740
 1745
 1750
 1755
 1760
 1765
 1770
 1775
 1780
 1785
 1790
 1795
 1800
 1805
 1810
 1815
 1820
 1825
 1830
 1835
 1840
 1845
 1850
 1855
 1860
 1865
 1870
 1875
 1880
 1885
 1890
 1895
 1900
 1905
 1910
 1915
 1920
 1925
 1930
 1935
 1940
 1945
 1950
 1955
 1960
 1965
 1970
 1975
 1980
 1985
 1990
 1995
 2000
 2005
 2010
 2015
 2020
 2025
 2030
 2035
 2040
 2045
 2050
 2055
 2060
 2065
 2070
 2075
 2080
 2085
 2090
 2095
 2100
 2105
 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250
 2255
 2260
 2265
 2270
 2275
 2280
 2285
 2290
 2295
 2300
 2305
 2310
 2315
 2320
 2325
 2330
 2335
 2340
 2345
 2350
 2355
 2360
 2365
 2370
 2375
 2380
 2385
 2390
 2395
 2400
 2405
 2410
 2415
 2420
 2425
 2430
 2435
 2440
 2445
 2450
 2455
 2460
 2465
 2470
 2475
 2480
 2485
 2490
 2495
 2500
 2505
 2510
 2515
 2520
 2525
 2530
 2535
 2540
 2545
 2550
 2555
 2560
 2565
 2570
 2575
 2580
 2585
 2590
 2595
 2600
 2605
 2610
 2615
 2620
 2625
 2630
 2635
 2640
 2645
 2650
 2655
 2660
 2665
 2670
 2675
 2680
 2685
 2690
 2695
 2700
 2705
 2710
 2715
 2720
 2725
 2730
 2735
 2740
 2745
 2750
 2755
 2760
 2765
 2770
 2775
 2780
 2785
 2790
 2795
 2800
 2805
 2810
 2815
 2820
 2825
 2830
 2835
 2840
 2845
 2850
 2855
 2860
 2865
 2870
 2875
 2880
 2885
 2890
 2895
 2900
 2905
 2910
 2915
 2920
 2925
 2930
 2935
 2940
 2945
 2950
 2955
 2960
 2965
 2970
 2975
 2980
 2985
 2990
 2995
 3000
 3005
 3010
 3015
 3020
 3025
 3030
 3035
 3040
 3045
 3050
 3055
 3060
 3065
 3070
 3075
 3080
 3085
 3090
 3095
 3100
 3105
 3110
 3115
 3120
 3125
 3130
 3135
 3140
 3145
 3150
 3155
 3160
 3165
 3170
 3175
 3180
 3185
 3190
 3195
 3200
 3205
 3210
 3215
 3220
 3225
 3230
 3235
 3240
 3245
 3250
 3255
 3260
 3265
 3270
 3275
 3280
 3285
 32

1036 GATTTCAGTGAAGAGATGTGGCAGCAAGAGATCCCCGGGACCGAATCCAGTTTCAGG 1095

[illegible]

| | | |
|-----|---|-----|
| D0 | CCCTACGCAAAATGGAAAGTGTGCCCAGCCACACCTTGGAAAGGAGGAGAAATATCATCATCTATGACCTC | 423 |
| 304 | CCCTACGCAAAATGGAAAGTGTGCCCAGCCACACCTTGGAAAGGAGGAGAAATATCATCATCTATGACCTC | 423 |
| QY | 1156 CCTACAGGGAGTGGAAAAACAGAGTGGCTGTTTACATATGCCAAGGATCAGCTATATACAG 1215 | |
| 424 | CCCTACGCAAAATGGAAAGTGTGCCCAGCCACACCTTGGAAAGGAGGAGAAATATCATCATCTATGACCTC | 423 |

[illegible][illegible]

| QY | 1335 | ATTAACT-GGTGATACCCAACTGAAAATATATATTTCCAGAAAGTTGTCAATGC-1386 |
|-----|------|---|
| 00 | | |
| 04 | | |
| 08 | | |
| 12 | | |
| 16 | | |
| 20 | | |
| 24 | | |
| 28 | | |
| 32 | | |
| 36 | | |
| 40 | | |
| 44 | | |
| 48 | | |
| 52 | | |
| 56 | | |
| 60 | | |
| 64 | | |
| 68 | | |
| 72 | | |
| 76 | | |
| 80 | | |
| 84 | | |
| 88 | | |
| 92 | | |
| 96 | | |
| 100 | | |
| 104 | | |
| 108 | | |
| 112 | | |
| 116 | | |
| 120 | | |
| 124 | | |
| 128 | | |
| 132 | | |
| 136 | | |
| 140 | | |
| 144 | | |
| 148 | | |
| 152 | | |
| 156 | | |
| 160 | | |
| 164 | | |
| 168 | | |
| 172 | | |
| 176 | | |
| 180 | | |
| 184 | | |
| 188 | | |
| 192 | | |
| 196 | | |
| 200 | | |
| 204 | | |
| 208 | | |
| 212 | | |
| 216 | | |
| 220 | | |
| 224 | | |
| 228 | | |
| 232 | | |
| 236 | | |
| 240 | | |
| 244 | | |
| 248 | | |
| 252 | | |
| 256 | | |
| 260 | | |
| 264 | | |
| 268 | | |
| 272 | | |
| 276 | | |
| 280 | | |
| 284 | | |
| 288 | | |
| 292 | | |
| 296 | | |
| 300 | | |
| 304 | | |
| 308 | | |
| 312 | | |
| 316 | | |
| 320 | | |
| 324 | | |
| 328 | | |
| 332 | | |
| 336 | | |
| 340 | | |
| 344 | | |
| 348 | | |
| 352 | | |
| 356 | | |
| 360 | | |
| 364 | | |
| 368 | | |
| 372 | | |
| 376 | | |
| 380 | | |
| 384 | | |
| 388 | | |
| 392 | | |
| 396 | | |
| 400 | | |
| 404 | | |
| 408 | | |
| 412 | | |
| 416 | | |
| 420 | | |
| 424 | | |
| 428 | | |
| 432 | | |
| 436 | | |
| 440 | | |
| 444 | | |
| 448 | | |
| 452 | | |
| 456 | | |
| 460 | | |
| 464 | | |
| 468 | | |
| 472 | | |
| 476 | | |
| 480 | | |
| 484 | | |
| 488 | | |
| 492 | | |
| 496 | | |
| 500 | | |
| 504 | | |

RESULT 13

| | | | |
|----|-----|---|-----|
| UD | 604 | ATTACCTGGGAGTAACCCACACGAAAAATATCATTTCCAGAGAGTGTCCAGATTC | 750 |
|----|-----|---|-----|

| | | | | | |
|------------|----------------------------------|--------|--------------|--------|----------------|
| LOCUS | BQ16075 | 672 bp | mRNA | linear | 17-MAY-2002 |
| DEFINITION | CMV-CT0275-191099-024-g09 CT0275 | | Homo sapiens | CDNA. | |
| ACCESSION | BQ16075 | | | | mRNA sequence. |

| | | |
|----------|--------------|-------------|
| VERSION | BQ316075.1 | GI:20921844 |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |

REFERENCE
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M. R.,
Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 672)

Nagai, M. A., da Silva, W. Jr., Zaço, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsumura, A., Bata, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and

Simpson, A.J.
Shotgun sequencing of the human transcriptome with 454 expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM3&l2=CM3-CT0275-
191099-024-909&l3=1999-10-19&l4=1)

Seq primer: puc 18 forward
High quality sequence start: 50
High quality sequence stop: 671.
Location/Qualifiers

FEATURES

source

1. 672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0275"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site: 1; Smal; Site: 2;
Sma1: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 165 a 158 c 120 g 229 t

ORIGIN

Query Match 17.8%; Score 599.2; DB 14; Length 672;
Best Local Similarity 96.7%; Pred. No. 6.1e-118;
Matches 623; Conservative 0; Mismatches 18; Indels 3; Gaps 1;

2121 TGGATTTGATGGATGAGATGATGATTTAAAGAACTTGAAGCTGATGA 2180
672 TGGATTTGATGGATGAGATGAGATGATTTAAAGAACTTGAAGCTGAG 613
2181 AACAGATAGATTTCTCATGACTTATTTTGAACAATAAATGTTGAAGCTGAG 2240
612 AACAGATAGATTTCTCATGACTTATTTTGAACAATAAATGTTGAAGCTGAG 553
2241 TGAAGCCAGATATGAAATGAAAAGCTGACCAATTAAGAAATACCAATGACGTA 2300
552 TGAAGCCAGATATGAAATGAAAAGCTGACCAATTAAGAAATACCAATGACGTA 493
2301 ATATAGTACGATGAGATGAGATGAGATGATTTAAAGAACTTGAAGCTGAG 2360
492 ATATAGTACGATGAGATGAGATGAGATGATTTAAAGAACTTGAAGCTGAG 433
2361 ATATGCGCTTTCCAGTGGATTTACTGAAAATGAAAATTTGCTGAAGTAAGTCAAGC 2420
432 ATATGCGCTTTCCAGTGGATTTACTGAAAATGAAAATTTGCTGAAGTAAGTCAAGC 373
2421 CCACCATCTGATTTGAGCTGGACACAGACAGTGTCAACCCATGACATTAATGACA 2480
372 CCACCATCTGATTTGAGCTGGACACAGACAGTGTCAACCCATGACATTAATGACA 313
2481 AAAGAGATCATTAAGTAAATTTGCACTGGAAATATCATTCGCTATGTAACCAAGT 2540
312 AAAGAGATCATTAAGTAAATTTGCACTGGAAATATCATTCGCTATGTAACCAAGT 253
2541 GGCAGAGAAGATCTGATTTAAAGATGTAACATTTATCGCTATCTCTGCTGAC 2600
252 GGCAGAGAAGATCTGATTTAAAGATGTAACATTTATTCGCTATCTCTGCTGAC 193
2601 CAATGAATAGCCATGCTCCAGGCCGCTGCTGAGCCAGACAGTGAATAGTACCTACAGT 2660
192 CCATGAATAGCCATGCTCCAGGCCGCTGCTGAGCCAGACAGTGAATAGTACCTACAGT 133
2661 CCGGTTGCTGCAAGTGGTTCAGAGTTATCGAACAATGACAGTGAATTTTTCGAGAA 2720
132 CCGGTTGCTGCAAGTGGTTCAGAGTTATCGAACAATGACAGTGAATTTTTCGAGAA 73
2721 GAAGATGATGATTAAGTATACATTTGTTCAAAATATGAAGC 2764

Db 72 GAAGATGATGATTAAGTATACATTTGTTCAAAATATGAAGC 32

RESULT 14

LOCUS BQ316108 672 bp mRNA linear EST 17-MAY-2002
DEFINITION CM3-CT0275-221099-024-909 CT0275 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ316108
VERSION BQ316108.1 GI:20921877
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 672)
Nas Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Dagal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.G.

Shotgun sequencing of the human transcriptome with OEF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM3&l2=CM3-CT0275-
191099-024-909&l3=1999-10-22&l4=1)

Seq primer: puc 18 forward
High quality sequence start: 50
High quality sequence stop: 671.
Location/Qualifiers

FEATURES

source

1. 672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0275"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site: 1; Smal; Site: 2;
Sma1: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 165 a 158 c 120 g 229 t

ORIGIN

Query Match 17.8%; Score 599.2; DB 14; Length 672;
Best Local Similarity 96.7%; Pred. No. 6.1e-118;
Matches 623; Conservative 0; Mismatches 18; Indels 3; Gaps 1;

2121 TGGATTTGATGGATGAGATGAGATGATTTAAAGAACTTGAAGCTGATGA 2180
672 TGGATTTGATGGATGAGATGAGATGATTTAAAGAACTTGAAGCTGAG 613
2181 AACAGATAGATTTCTCATGACTTATTTTGAACAATAAATGTTGAAGCTGAG 2240
612 AACAGATAGATTTCTCATGACTTATTTTGAACAATAAATGTTGAAGCTGAG 553
2241 TGAAGCCAGATATGAAATGAAAAGCTGACCAATTAAGAAATACCATATTAAGCA 2300
552 TGAAGCCAGATATGAAATGAAAAGCTGACCAATTAAGAAATACCATATTAAGCA 493

/organism-"Homo sapiens"

DB 1 ACAAACAGTATCCCTCTCCAGATCTGGGAGTAACAGCTTACCTGCTGGAG 60
QY 1652 GGGCCAGAGAGAGCAAGTGAAGACATTTTAAATATGTGCAATCTTGATG 1711
DB 61 GGGCCAGAGAGAGCAAGTGAAGACATTTTAAATATGTGCAATCTTGATG 120
QY 1712 CATTTACTATTTAACTGTTAAAGAAACCTTGATCACTGAAACCAATACAGAGAT 1771
DB 121 CATTTACTATTTAACTGTTAAAGAAACCTTGATCACTGAAACCAATACAGAGAT 179
QY 1772 CATGCAAGAGTTGGCATTCAGATGCAACAGAGAGATTCATTTAAAGAGAACTTG 1831
DB 180 CATGCAAGAGTTGGCATTCAGATGCAACAGAGAGATTCATTTAAAGAGAACTTG 239
QY 1832 TAGAAATATGACAGAGATTCAACTTATGTCAATGATCCAAATGTCATTTTGA 1890
DB 240 TAGAAATATGACAGAGATTCAACTTATGTCAATGATCCAAATGTCATTTTGA 298

RESULT 2

US-08-143-576-6
Sequence 6, Application US/08143576
Patent No. 5643761

GENERAL INFORMATION:

APPLICANT: Fisher, Paul B.
TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRUCTURE
TITLE OF INVENTION: CDNA LIBRARY AND USES OF THE GENERATED LIBRARY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White, c/o Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/143,576

FILING DATE: 25-OCT-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 43563/JPM/AMC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UT

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 193 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-143-576-6

Query Match

Best Local Similarity 3.7%; Score 124.6; DB 1; Length 193;
Pred. No. 4.8e-23;

Matches 138; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 3007 CTGCAAAAGAGTGTCCGACTATCAATAATGATGTAATCACTGCAATGTCGCGAG 3066
DB 1 CTGCAAAAGAGTGTCCGACTAT-AAAATAATGATGTAATCACTGCAATGTCGCGAG 59

QY 3067 GCTGGGAGACATGATGTCGACAAAGCGTTAGATTGCTGTCTCAAAATACAT 3126
DB 60 GCTGGGAGACATGATGTCGACAAAGCGTTAGATTGCTGTCTCAAAATACAT 119
QY 3127 TTTGTAGTGTTCATAAAATTA 3149
DB 120 TTTGTAGTGTTCATAAAATTA 142

RESULT 3

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.
APPLICANT: SCHIEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)863-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZ9pt-Fls

US-08-232-463-14

Query Match

Best Local Similarity 2.1%; Score 69.4; DB 1; Length 7218;
Pred. No. 6e-08; Mismatches 146; Indels 0; Gaps 0;

Matches 19; Conservative 230; Mismatches 146; Indels 0; Gaps 0;

QY 1876 ATGTGAGATTGCAACTACACCTATGACACATGCGCCATTCAATGAAAGAAAGCT 1935

DB 1451 ATAGAGAGATTGTGACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392

QY 1936 GCAAAAGAGAGATGCAAGAGAGTGTGTCGACAGACATTTGAGAGAGAGATGAG 1995

DB 1391 RRR 1332

QY 1996 GCGCAGCAATTAATGACAGATGATAGATGCTACTCACTCTCAATTTTC 2055

DB 1331 RRR 1272

Db 1610 TTTTGATTGGCCAGTGAATTCAGAGAAATATGTCATTCGTTATGCGCGTAAAGAGCTGT 1669
QY 2646 TGAGAGC 2652
Db 1670 AGGAAAC 1676

RESULT 10

US-09-058-489-17
; Sequence 17, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 4416
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-17

Query Match 1.4%; Score 47.8; DB 3; Length 4416;
Best Local Similarity 53.5%; Pred. No. 0.018;
Matches 100; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 2466 GACACAGATGAACAAAAAGCTCATTAATAATTCGACCTGSAANAATATCTGCT 2525
Db 1490 GTCACAGAGATGAGAGAGAGGCGCCCTTCACAGTTCCGCTGCAAGAAATGCCAATCT 1549
QY 2526 TATCCGTACACAGTGGCACAAGAGCTCGATATTAAAGATTTAAATTTGTTATCCG 2585
Db 1550 AGGGCTACAGCTGTGGCAGCAGAGACTAGACATTTCAAATGAGAGATTTTAA 1609
QY 2586 TTAATGCTGTGTGACCAATGAATATAGCCATGTCACAGCGCGCTGTGCAATAGAGCTGA 2645
Db 1610 TTTTGATTGGCCAGTGAATTCAGAGAAATATGTCATTCGTTATGCGCGTAAAGAGCTGT 1669
QY 2646 TGAGAGC 2652
Db 1670 AGGAAAC 1676

RESULT 11

US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 1.4%; Score 47.8; DB 4; Length 5361;
Best Local Similarity 50.5%; Pred. No. 0.02;
Matches 143; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

QY 2026 ATAGATGCTATACATCTTGAACCTTTCTAATGAAGAAAGATTAAGAAATTTGCA 2085
Db 3544 ATATCTTCAGATTCTTAAGCAAGAACTGATATTAAAGATAAGAAATTTTCA 3603
QY 2086 GTCATGAAGATGATAGTATGATGAGGCTGTGATGATGATGATGATGATGATGAT 2145
Db 3604 CTAGTTGTTGAAGCACTTAAGCAATGATATGATGATGATGATGATGATGATGATGAT 3663
QY 2146 GAGGATGATTTAAGAAACCTTTGAACCTGATGAACAGATGATGATGATGATGATGAT 2204
Db 3664 TGAAGAAATATGAAAGAGGATTAATGAAGATGCTGTTGAATGAATGACATTAAGTAC 3723
QY 2205 -ATTTTGAAGAAATTAATGTTGAAGAGCTGCTGGAAGAAACCCAGATATGAAT 2262
Db 3724 AAACCTATTTGAAGAACTCAAGAGTTAAATGAAGTGAAGAGGATTTAATAAAGATATG 3783
QY 2263 GAAAGCTGACCAATTAAGAAATACCATATGAGCAATATA 2305
Db 3784 GAAATTTAAGAAATTAAGAAATTAAGCAATATATGAAAGATTTCA 3826

RESULT 12

US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 1.4%; Score 47.8; DB 4; Length 6152;
Best Local Similarity 50.5%; Pred. No. 0.02;
Matches 143; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

QY 2026 ATAGATGCTATACATCTTGAACCTTTCTAATGAAGAAAGATTAAGAAATTTGCA 2085
Db 3788 ATATCTTCAGATTCTTAAGCAAGAACTGATATTAAAGATAAGAAATTTTCA 3847
QY 2086 GTCATGAAGATGATAGTATGATGAGGCTGTGATGATGATGATGATGATGATGAT 2145
Db 3848 CTAGTTGTTGAAGCACTTAAGCAATGATATGATGATGATGATGATGATGATGATGAT 3907
QY 2146 GAGGATGATTTAAGAAACCTTTGAACCTGATGAACAGATGATGATGATGATGATGAT 2204
Db 3908 TTAAGAAATATGAAAGAGGATTAATGAAGATGCTGTTGAATGAATGACATTAAGTAC 3967
QY 2205 -ATTTTGAAGAAATTAATGTTGAAGAGCTGCTGGAAGAAACCCAGATATGAAT 2262
Db 3968 AAACCTATTTGAAGAACTCAAGAGTTAAATGAAGTGAAGAGGATTTAATAAAGATATG 4027


```

1 STREET: 225 Franklin Street Suite 3100
2 CITY: Boston
3 STATE: MA
4 COUNTRY: USA
5 ZIP: 02110-2804
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patent Release #1.0, Version #1.30
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: PCT/US95/16930
12 FILING DATE: 27-DEC-1995
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: US 08/375,300
15 FILING DATE: 20-JAN-1995
16 ATTORNEY/AGENT INFORMATION:
17 NAME: Fasse, J. Peter
18 REGISTRATION NUMBER: 32,983
19 REFERENCE/DOCKET NUMBER: 04020/046W01
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: (617)542-5070
22 TELEFAX: (617)542-8906
23 TELEX: 200154
24 INFORMATION FOR SEQ ID NO: 3:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 2295 base pairs
27 TYPE: nucleic acid
28 STRANDEDNESS: single
29 TOPOLOGY: linear
30 MOLECULE TYPE: DNA
31 PCT-US95-16930-3

```

| | | | | |
|-----------------------|-----------------|------------------|-----------------|--------------|
| Query Match | 1.4% | Score 47.2; | DB 5; | Length 2295; |
| Best Local Similarity | 51.4%; | Pred. No. 0.019; | Mismatches 103; | Gaps 0; |
| Matches 109; | Conservative 0; | | | |

```

QY      2057 ATAAAGACAGACAATAAAGATTGGCAGCGCATGTACGATATATAGTGAAGGTCGTG 2116
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1718 ATGACGATGACACATACGCATGATGATGATGATGAGAGAACAGCCGATGATGATCATG 1777
QY      2117 ATGATGACTATTGTGATGGTGATGAAGAATGACGATGATTTAAAGAAAGCTTGGAAACTGG 2176
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1778 AAGATGATGATGATGATGAGGATGATGATGATGAAGAAGAACAGACAGCGGATCTGATTTGG 1837
QY      2177 ATGAACACGATATGATTTCTCATGCACTTAATTTTTTGAAGAACATAAATATTTGAAAAGGC 2236
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1838 AGTATGCGTGCTGATCTTGACGCCAGACAGAGATATTGAAATGAACGATATATGAAGAGCT 1897
QY      2237 TGGCTGAAAAACCACGATATGAAATGAAAAG 2268
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1898 ACGAGAGAAACTTAAAGGATGAGAGAAAGG 1929

```

Search completed: May 16, 2003, 13:24:45
Job time : 294 secs

PN MO200164707-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06960.
 XX
 PR 29-FEB-2000; 2000US-0515363.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Fisher PB, Kang D, GopalKrishnan RV;
 XX
 DR WPI: 2001-565494/63.
 DR P-PSDB: AAEI0155.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity
 XX
 PS Claim 1: Page 16-18; 152pp; English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC antileukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human Mda-5 cDNA.
 XX
 SO Sequence 3365 BP; 1138 A; 644 C; 753 G; 830 T; 0 other.
 Query Match 100.0%; Score 3365; DB 22; Length 3365;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 421 AGAAGCGGAGCCCTCTGCGCCCGGCTACATGAACCTGACCTACGGACTTCCCTCT 480
 Qy 481 CCATGGTTTGAGAACGCTCATGATGAATATCTCAACTGTGTAACCTCTTACGCCACT 540
 Db 481 CCATGGTTTGAGAACGCTCATGATGAATATCTCAACTGTGTAACCTCTTACGCCACT 540
 Qy 541 CTGCTGAGACAGCTTCTAGTATGACAGCTCTTGATTAAGTGCTAGGAGAGAGACTGTTG 600
 Db 541 CTGCTGAGACAGCTTCTAGTATGACAGCTCTTGATTAAGTGCTAGGAGAGAGACTGTTG 600
 Qy 601 ACAATTGAAAGACAGAAACCGGATGCTGTCGAGAAACCAATGGAATATCAGGTGTA 660
 Db 601 ACAATTGAAAGACAGAAACCGGATGCTGTCGAGAAACCAATGGAATATCAGGTGTA 660
 Qy 661 AGAGAGCTACTAAAAAGATTGTGAGAAAGAAACCTGTTCTGCAATTTCTGAATGTT 720
 Db 661 AGAGAGCTACTAAAAAGATTGTGAGAAAGAAACCTGTTCTGCAATTTCTGAATGTT 720
 Qy 721 CTTCGTCAAACAGAAACCAATGCTTGTCCAAAGATTACAGGCTCTGATTTCTACAA 780
 Db 721 CTTCGTCAAACAGAAACCAATGCTTGTCCAAAGATTACAGGCTCTGATTTCTACAA 780
 Qy 781 ACCAATGAGAGATGAGATTAATACAGAGTTGATGCTTCAAGTGAAGAGCAACTT 840
 Db 781 ACCAATGAGAGATGAGATTAATACAGAGTTGATGCTTCAAGTGAAGAGCAACTT 840
 Qy 841 CTTTCACCCACAGTTCAGCCAAATCTGAGAAAGAGTCTGGGATGAGAAATATCTCA 900
 Db 841 CTTTCACCCACAGTTCAGCCAAATCTGAGAAAGAGTCTGGGATGAGAAATATCTCA 900
 Qy 901 TCAGATATATCTTTTGCAATTTCTGCTAGTTTCAAGTACAGACACAAAGTTTGGCAGAA 960
 Db 901 TCAGATATATCTTTTGCAATTTCTGCTAGTTTCAAGTACAGACACAAAGTTTGGCAGAA 960
 Qy 961 GGAATGTCAGTGTGTTGATGAAGAAAGCTTGAGACATAACAGCAATGGGAGTATCA 1020
 Db 961 GGAATGTCAGTGTGTTGATGAAGAAAGCTTGAGACATAACAGCAATGGGAGTATCA 1020
 Qy 1021 GGCACCATGGAAGATGATTCAGATGAGAAATGTGGCAGCAAGCATCCCGAGCCA 1080
 Db 1021 GGCACCATGGAAGATGATTCAGATGAGAAATGTGGCAGCAAGCATCCCGAGCCA 1080
 Qy 1081 GAATCCACGCTAGGCTTTACCAATGGAATGTGGCCAGCCAGCTTGGAAAGGAAAGAT 1140
 Db 1081 GAATCCACGCTAGGCTTTACCAATGGAATGTGGCCAGCCAGCTTGGAAAGGAAAGAT 1140
 Qy 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACCAAGTGGCTGTATTAATGCCAAAG 1200
 Db 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACCAAGTGGCTGTATTAATGCCAAAG 1200
 Qy 1201 GATCACTTAGACAGAGAAAGAGCATCTGAGCCTGGAAAAAGTTATGTTCTGTCAAT 1260
 Db 1201 GATCACTTAGACAGAGAAAGAGCATCTGAGCCTGGAAAAAGTTATGTTCTGTCAAT 1260
 Qy 1261 AAGTACTGCTAGTGAACAGCTCTTCGCAAGAGGATTCACCAATTTTGAAGAAATGG 1320
 Db 1261 AAGTACTGCTAGTGAACAGCTCTTCGCAAGAGGATTCACCAATTTTGAAGAAATGG 1320
 Qy 1321 TATCGTATTGATTAAGTGAAGTGAATACCAACATGAAATATATATTTCCAGAGTGTG 1380
 Db 1321 TATCGTATTGATTAAGTGAAGTGAATACCAACATGAAATATATATTTCCAGAGTGTG 1380
 Qy 1381 AAGTCTGTGATTAATATAGTAGACAGCTCAATCTTGAAGAACTCCCTTTAACTTGG 1440
 Db 1381 AAGTCTGTGATTAATATAGTAGACAGCTCAATCTTGAAGAACTCCCTTTAACTTGG 1440
 Qy 1441 GAAAAATGAGAAAGATGCTGGGTTCATATGTCACACTTTTCCCATATATATGATGAA 1500
 Db 1441 GAAAAATGAGAAAGATGCTGGGTTCATATGTCACACTTTTCCCATATATATGATGAA 1500
 Qy 1501 TGTCTATCACCAACAAAGAGAGAGTATATATTAACATCATGAGGCAATATTTGATTCAG 1560
 Db 1501 TGTCTATCACCAACAAAGAGAGTATATATTAACATCATGAGGCAATATTTGATTCAG 1560


```

FT      /tag= a
ET      /product= "Human RH16"
XX      WO200185955-A1.
XX      15-NOV-2001.
XX      11-MAY-2001: 2001WO-FR01441.
XX      11-MAY-2000: 2000FR-0006030.
XX      (ISTA-) ISTAC.
XX      (INSP) INST PASTEUR LILLE.
XX      Bahr G, Cocude C, Capron A;
XX      WPI: 2002-082898/11.
XX      P-PSDB: AAM47798.
XX      New polypeptide, useful for treating and diagnosing cancer or
XX      inflammation, and drug screening, comprises a human polynucleotide
XX      homologous to RNA helicase.
XX      Claim 7, Page 85-89; 11pp; French.
XX      The present sequence is the coding sequence for human RH16. RH16 is a
XX      116kDa protein and has homology to RNA helicases (DEX1 box). RH16 and
XX      its coding sequence are useful for treating cancer, acute or chronic
XX      infections (especially by HIV or hepatitis B or C); inherited genetic
XX      diseases (auto)immune diseases (particularly rheumatism, arthritis,
XX      arteriosclerosis, osteoporosis and diabetes), but many others listed) and
XX      to prevent graft rejection. RH16 and its coding sequence are also useful
XX      for inducing, or increasing, the immune response to a vaccine.
XX      Sequence 3372 BP; 1157 A; 637 C; 748 G; 830 T; 0 other:
SQ
Query Match      99.2%; Score 3338.8; DB 24; Length 3372;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3343; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

DB      481  |||||||
DB      CGCTCATGATGAAATATCTCCAACTGCTCAACCTCTTCAGCCACTGCTGGTGGAAGCT 540
OY      555  TCTAGTAGAGACGCTTGATGAAGTGCATGGAGAGAGACTGTGACAAATTAAGACAG 614
DB      541  |||||||
DB      TCTAGTAGAGACGCTTGATGAAGTGCATGGAGAGAGACTGTGACAAATTAAGACAG 600
OY      615  AACCAGATGCTGCTGCGAGAAACAAATGAAATGAAATGAGTGTAGAGAGCTACTAAA 674
DB      601  AACCAGATGCTGCTGCGAGAAACAAATGAAATGAAATGAGTGTAGAGAGCTACTAAA 660
OY      675  AAGCATTTGTGAGAAAGAAAGAAAGTGTGCTGCAATTTGTAATGTTCTCTTAACAGG 734
DB      661  AAGCATTTGTGAGAAAGAAAGTGTGCTGCAATTTGTAATGTTCTCTTAACAGG 720
OY      735  AACCAGATGCTGCTGCGAGAAACAAATGAAATGAAATGAGTGTAGAGAGCTACTAAA 794
DB      721  AACCAGATGCTGCTGCGAGAAACAAATGAAATGAAATGAGTGTAGAGAGCTACTAAA 780
OY      795  TGAGAAATTTATCACAAGTTGATGCTCTCAAGTGGAGAGCACTTCTTCAACAGAGT 854
DB      781  TGAGAAATTTATCACAAGTTGATGCTCTCAAGTGGAGAGCACTTCTTCAACAGAGT 840
OY      855  TCAGCAAAATCTGAGAGAGAGGCTGTGGGCTATGAGAAATACATCATCATCTTT 914
DB      841  TCAGCAAAATCTGAGAGAGAGGCTGTGGGCTATGAGAAATACATCATCATCTTT 900
OY      915  TGCAATTTCTCTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 974
DB      901  TGCAATTTCTCTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 960
OY      975  CTAGATGAAAGTCTTGAGCAATTAACAGCAATGAGGAGTATGATGATGATGATGATG 1034
DB      961  CTAGATGAAAGTCTTGAGCAATTAACAGCAATGAGGAGTATGATGATGATGATGATG 1020
OY      1035  TGATTCATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1094
DB      1021  TGATTCATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1080
OY      1095  GCCTTACCAAAATGGAAGTTGGCCAGCCAGCCTTGGAGAGAGAAATATCATCATCTGCT 1154
DB      1081  GCCTTACCAAAATGGAAGTTGGCCAGCCAGCCTTGGAGAGAGAAATATCATCATCTGCT 1140
OY      1155  CCCTACAGAGAGTGGAAAAACAGAGTGGCTGTATTACTTCCCAAGATCACTTAAGACA 1214
DB      1141  CCCTACAGAGAGTGGAAAAACAGAGTGGCTGTATTACTTCCCAAGATCACTTAAGACA 1200
OY      1215  GAAGAAAAAAGCATGCTGAGCTGGAAAAATGATGTTGTCATTAAGATGATGATGATG 1274
DB      1201  GAAGAAAAAAGCATGCTGAGCTGGAAAAATGATGTTGTCATTAAGATGATGATGATG 1260
OY      1275  TGAACAGCTCTTCGCAAGAGATTCACACATTTTGAAGAAATGATGATGATGATGATG 1334
DB      1261  TGAACAGCTCTTCGCAAGAGATTCACACATTTTGAAGAAATGATGATGATGATGATG 1320
OY      1335  ATTAAGTGTGATATCCCACTGAAAAATATCATTTTCCAGAAAGTGTGCAAGTCTGTAT 1394
DB      1321  ATTAAGTGTGATATCCCACTGAAAAATATCATTTTCCAGAAAGTGTGCAAGTCTGTAT 1380
OY      1395  TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1454
DB      1381  TATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
OY      1455  TCGTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1514
DB      1441  TCGTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
OY      1515  CAAGAAACAGTGTATTAATTAATCATGATGAGCAATTAATTTGATGAGAAATGAAACAA 1574
DB      1501  CAAGAAACAGTGTATTAATTAATCATGATGAGCAATTAATTTGATGAGAAATGAAACAA 1560
OY      1575  TAGATCAAGAAAGAAACCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1634

```

Db 1561 TAGACTCAGAGAAAGAAACCAACAGTATGTCCTCTCAGACTACTGTTACTAACAGC 1620
 QY 1635 TTGACCTGGTGTGGAGGGGCGACGAGAGCAAGCGTAGACACAGCTTTAAACCT 1694
 Db 1621 TTGACCTGGTGTGGAGGGGCGACGAGAGCAAGCGTAGACACAGCTTTAAACCT 1680
 QY 1695 ATGTGGCAATCTTGATGATCTTACTATTTAAACTGTTAAGAAACCTTATGACCTGAA 1754
 Db 1681 ATGTGGCAATCTTGATGATCTTACTATTTAAACTGTTAAGAAACCTTATGACCTGAA 1740
 QY 1755 AAACCAATACAGAGAGCAGTCAAGAGAGTTGGCATTGCGATGCAACCTGAGACATCC 1814
 Db 1741 AAACCAATACAGAGAGCAGTCAAGAGAGTTGGCATTGCGATGCAACCTGAGACATCC 1800
 QY 1815 ATTTAAGAGAACTCTAGAAATTAATGACAGAGATTCAACTATTTGTAAATGAGTCC 1874
 Db 1801 ATTTAAGAGAACTCTAGAAATTAATGACAGAGATTCAACTATTTGTAAATGAGTCC 1860
 QY 1875 AATGTCAGATTTTGAACCTCAGCCTATGAAACATGGCCATTCAATG AAAAAAGC 1934
 Db 1861 AATGTCAGATTTTGAACCTCAGCCTATGAAACATGGCCATTCAATG AAAAAAGC 1920
 QY 1935 TCGAAAAAGAGAAATCGCAAGACGCTTTTGTGCAAGACATTTGAGTGTACAAATGA 1994
 Db 1921 TCGAAAAAGAGAAATCGCAAGACGCTTTTGTGCAAGACATTTGAGTGTACAAATGA 1980
 QY 1995 GGGCCCTACAAATTAATGACCAATTTGAATGATAGATGCTATACGATTTGAACCTT 2054
 Db 1981 GGGCCCTACAAATTAATGACCAATTTGAATGATAGATGCTATACGATTTGAACCTT 2040
 QY 2055 CTATTAATGAGAGAAAGATTAAGAGTTGGCAGTCATGAGATGATAGTTAGAGGCTGG 2114
 Db 2041 CTATTAATGAGAGAAAGATTAAGAGTTGGCAGTCATGAGATGATAGTTAGAGGCTGG 2100
 QY 2115 TGATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2174
 Db 2101 TGATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2175 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2234
 Db 2161 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 QY 2235 GCTGGCTGAGAAACCCAGATATGAAAAAGCTGACCAATTAAGATACCAATAT 2294
 Db 2221 GCTGGCTGAGAAACCCAGATATGAAAAAGCTGACCAATTAAGATACCAATAT 2280
 QY 2295 GAGACCAATATAGTACGATGAGAAATCAGACAGAGAAATATCTTTACATTAACAGGACA 2354
 Db 2281 GAGACCAATATAGTACGATGAGAAATCAGACAGAGAAATATCTTTACATTAACAGGACA 2340
 QY 2355 GAGTGCATATGCGCTTCCAGAGTGAATGATGATGATGATGATGATGATGATGATGAT 2414
 Db 2341 GAGTGCATATGCGCTTCCAGAGTGAATGATGATGATGATGATGATGATGATGATGAT 2400
 QY 2415 CAAAGCCACCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2474
 Db 2401 CAAAGCCACCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 QY 2475 TGAACAAAAAGAGTCTAGTAAATTTGCGACCTGAGAAATATCATCTGTATATCCCTAC 2534
 Db 2461 TGAACAAAAAGAGTCTAGTAAATTTGCGACCTGAGAAATATCATCTGTATATCCCTAC 2520
 QY 2535 CACAGTGGCAGAAAGAGTGTGATATTAAGAAATGATGATGATGATGATGATGATGATGAT 2594
 Db 2521 CACAGTGGCAGAAAGAGTGTGATATTAAGAAATGATGATGATGATGATGATGATGATGAT 2580
 QY 2595 CGTACCAATGAAATAGCAGTGTGCGAGCCGCTGGTGGCGAGAGCTATGAGAGCAG 2654
 Db 2581 CGTACCAATGAAATAGCAGTGTGCGAGCCGCTGGTGGCGAGAGCTATGAGAGCAG 2640
 QY 2655 CTAGCTCTGCTGCTCAGAGTGTTCAGAGTATGAGATGAGACATGAGACATTAATGATTT 2714
 Db 2641 CTAGCTCTGCTGCTCAGAGTGTTCAGAGTATGAGATGAGACATGAGACATTAATGATTT 2700

QY 2715 CCGAGAGAGATGATGATTAAGCTATACATTTGCTGTCAAAATATGAAACCAACAGAGTA 2774
 Db 2701 CCGAGAGAGATGATGATTAAGCTATACATTTGCTGTCAAAATATGAAACCAACAGAGTA 2760
 QY 2775 TCGCTATAGATTTTGAATTAACAGATGCAAGATTAATGAAAAAGAAATGAAACCA 2834
 Db 2761 TCGCTATAGATTTTGAATTAACAGATGCAAGATTAATGAAAAAGAAATGAAACCA 2820
 QY 2835 GAGAAATATTGCGAAGCATTAACAGATTAACCATCAGTAAATGATTTCTTTG AAAAA 2894
 Db 2821 GAGAAATATTGCGAAGCATTAACAGATTAACCATCAGTAAATGATTTCTTTG AAAAA 2880
 QY 2895 CTGCACTGTGAGCCCTGCTTGGGAAAGATTCATGATTAATGAAAAATGATACAGCT 2954
 Db 2881 CTGCACTGTGAGCCCTGCTTGGGAAAGATTCATGATTAATGAAAAATGATACAGCT 2940
 QY 2955 CATATGACCCGAGAAATTCAGAACTTTACATTTGTAAGAGAAACAAAGCAATGAAAA 3014
 Db 2941 CATATGACCCGAGAAATTCAGAACTTTACATTTGTAAGAGAAACAAAGCAATGAAAA 3000
 QY 3015 GAAGTGTGCGACATTCAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 3074
 Db 3001 GAAGTGTGCGACATTCAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 3060
 QY 3075 AACCAATGATGGCGACAAAGGCTTGAATTTGCTGTCGCAAAATAGGAAATTTTGTAGT 3134
 Db 3061 AACCAATGATGGCGACAAAGGCTTGAATTTGCTGTCGCAAAATAGGAAATTTTGTAGT 3120
 QY 3135 GCTTTCAAAATTAATTAACAAAGAAACATACAAAAAGTGGTAGATTAATGATGATGAT 3194
 Db 3121 GCTTTCAAAATTAATTAACAAAGAAACATACAAAAAGTGGTAGATTAATGATGATGAT 3180
 QY 3195 ATTTCCTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3254
 Db 3181 ATTTCCTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 QY 3255 TGAAGATCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3314
 Db 3241 TGAAGATCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 QY 3315 ATTATGCTACAGAACTGACATTAAGAAATCAATTAATGATGATGATGATGATGATGAT 3364
 Db 3301 ATTATGCTACAGAACTGACATTAAGAAATCAATTAATGATGATGATGATGATGATGAT 3350
 Db 3301 ATTATGCTACAGAACTGACATTAAGAAATCAATTAATGATGATGATGATGATGATGAT 3350
 RESULT 3
 AAD11170
 ID AAD11170 standard; DNA; 3131 BP.
 AC AAD11170;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE Human melanoma differentiation associated-5 protein-related DNA.
 XX
 KW Human melanoma differentiation associated gene; Mda-5; interferon; IFN;
 KW RNA helicase motif; caspase recruitment domain; CARD; the way; melanoma;
 KW neuroblastoma; astrocytoma; glioblastoma; multiforme; cancer; cervical;
 KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
 KW central nervous system; cytostatic; apoptosis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200164707-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06960.
 XX
 PR 29-FEB-2000; 2000US-051363.
 XX
 PA (UYCO) UNIV COLUMBIA NEM YORK.

QY 1789 ATTGCAGATGCAACGAGAGATCCATTTAAAGAAACTTCTGAAATATGCAACAG 1848
 DB 1618 ATTGCAGATGCAACGAGAGATCCATTTAAAGAAACTTCTGAAATATGCAACAG 1677
 QY 1849 ATTCAACTTATGTCAAATGAGTCCATGTGAGATTTTGGAACTCAATATGACAA 1908
 DB 1678 ATTCAAACTTATGTCAAATGAGTCCATGTGAGATTTTGGAACTCAATATGACAA 1737
 QY 1909 TGGGCGATTCAAATGAGAAAGAGTGGCAAAAAGAAATGCGAAATGAGTGTGT 1968
 DB 1738 TGGGCGATTCAAATGAGAAAGAGTGGCAAAAAGAAATGCGAAATGAGTGTGT 1797
 QY 1969 GCAGAACATTTAGAGAGATGACATGAGGCGCTACAAATTAATGACAAATGAAATGATA 2028
 DB 1798 GCAGAACATTTAGAGAGATGACATGAGGCGCTACAAATTAATGACAAATGAAATGATA 1857
 QY 2029 GAT--GCGTATACCTCTTGAACCTTCTATATGAGAGAAAGATTAATGATTGAG 2086
 DB 1858 GATCCGCGTATACCTCTTGAACCTTCTATATGAGAGAAAGATTAATGATTGAG 1917
 QY 2087 TCATGAAATGATGATGATGATGAGGCGTGTGATGATGATGATGATGATGATGATG 2146
 DB 1918 TCATGAAATGATGATGATGATGAGGCGTGTGATGATGATGATGATGATGATGATG 1977
 QY 2147 AGGATGATTTAAAGAACTTTGAAACCTGATGAAACAGATAGATTCTCTGACCTTAT 2206
 DB 1978 AGGATGATTTAAAGAACTTTGAAACCTGATGAAACAGATAGATTCTCTGACCTTAT 2037
 QY 2207 TTTTGGAAACATAAATGTTGAAAAAGGCTGGCTGAAACCCAGATAATGAAATGAAA 2266
 DB 2038 TTTTGGAAACATAAATGTTGAAAAAGGCTGGCTGAAACCCAGATAATGAAATGAAA 2097
 QY 2267 AGGTGCAAAATTAAGAAATACCATTAATGAGCAATATCTAGGAGTGTGATGATGAGC 2326
 DB 2098 AGGTGCAAAATTAAGAAATACCATTAATGAGCAATATCTAGGAGTGTGATGATGAGC 2157
 QY 2327 GAGGAATATCTTTACAAAAACGACGACAGATGATGAGGCTTCCCAATGATTAAGT 2386
 DB 2158 GAGGAATATCTTTACAAAAACGACGACAGATGATGAGGCTTCCCAATGATTAAGT 2217
 QY 2387 AAAATGAAAAATTTGCTGAATGAGAGTAAAGCCCATCTGATGAGGCTGACACA 2446
 DB 2218 AAAATGAAAAATTTGCTGAATGAGAGTAAAGCCCATCTGATGAGGCTGACACA 2277
 QY 2447 GAGGTGATGCAAAACCCATGACACAGATGAAAGCAAAAGAGCATTAGTAATTTGCA 2506
 DB 2278 GAGGTGATGCAAAACCCATGACACAGATGAAAGCAAAAGAGCATTAGTAATTTGCA 2337
 QY 2507 CTGGAATAATCAATCTGTTATGCTTACCAAGTGGCAGAGAGAGTCTTCAATTAAG 2566
 DB 2338 CTGGAATAATCAATCTGTTATGCTTACCAAGTGGCAGAGAGAGTCTTCAATTAAG 2397
 QY 2567 AATGTAACATTTGATCGCTTATGCTGCTGACCAATGAATAGCCATGCTCCAGGCC 2626
 DB 2398 AATGTAACATTTGATCGCTTATGCTGCTGACCAATGAATAGCCATGCTCCAGGCC 2457
 QY 2627 GTGGTGAGCAGAGCTGATGAGAGCACTAGCTCTGTTGCTGACAGTCTTTCAGAG 2686
 DB 2458 GTGGTGAGCAGAGCTGATGAGAGCACTAGCTCTGTTGCTGACAGTCTTTCAGAG 2517
 QY 2687 TTATGCAATGAGAGATGATGATTTCCGAGAGAAATGATGTAATCTATGCAAT 2746
 DB 2518 TTATGCAATGAGAGATGATGATTTCCGAGAGAAATGATGTAATCTATGCAAT 2577
 QY 2747 GTGTTAAAAATGAAACGAGAGAGTATGCTCATTAAGATTTTGAATTCAGATGAAA 2806
 DB 2578 GTGTTAAAAATGAAACGAGAGAGTATGCTCATTAAGATTTTGAATTCAGATGAAA 2637
 QY 2807 GTTATGAGAAAGAAATGAAAGCAAGAAATATTCAGAGCAATTAAGAAATTAAC 2866
 DB 2638 GTTATGAGAAAGAAATGAAAGCAAGAAATATTCAGAGCAATTAAGAAATTAAC 2697
 QY 2867 CATCACTAATTAATCTTCTTGGCAAAAGCTGACAGTGTGCTGAGCTTCTTCTGAGAGATA 2926

DB 2698 CATCACTAATTAATCTTCTTGGCAAAAGCTGACAGTGTGCTGAGCTGTCTGAGAGATA 2757
 QY 2927 TCCATGTAATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2986
 DB 2758 TCCATGTAATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2817
 QY 2987 TTGTAGAGAAACCAACGACTGCAAAAGAGTGGCGGACTATCAATTAATGATGAAA 3046
 DB 2818 TTGTAGAGAAACCAACGACTGCAAAAGAGTGGCGGACTATCAATTAATGATGAAA 2877
 QY 3047 TCATCTGCAATGTTGGCCAGGCTTGGGAGCAATGATGATGATGATGATGATGATGATG 3106
 DB 2878 TCATCTGCAATGTTGGCCAGGCTTGGGAGCAATGATGATGATGATGATGATGATGATG 2937
 QY 3107 CTGTCTCAAAATTAAGAAATTTTGTAGTGTGTTTCAAAAATTAATTAATCAACAAACAT 3166
 DB 2938 CTGTCTCAAAATTAAGAAATTTTGTAGTGTGTTTCAAAAATTAATTAATCAACAAACAT 2997
 QY 3167 ACAAAGAGGAGTACATTAATCAATTCATCCCAATCTTGACTATGAGATGATGATG 3226
 DB 2998 ACAAAGAGGAGTACATTAATCAATTCATCCCAATCTTGACTATGAGATGATGATG 3057
 QY 3227 TATTAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3286
 DB 3058 TATTAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3117
 QY 3287 ATTATATATGATTA 3300
 DB 3118 ATTATATATGATTA 3131

RESULT 4
 AAS40960
 ID AAS40960 standard; cDNA: 1967 BP.
 XX
 AC AAS40960;
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human enzyme polypeptide #176.
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW 11ase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 OS Homo sapiens.
 PN WO20015301-A2.
 XX
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01239.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198122.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225216.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0234401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246534.
 PR 08-NOV-2000; 2000US-0246535.
 PR 08-NOV-2000; 2000US-0246536.
 PR 08-NOV-2000; 2000US-0246537.
 PR 08-NOV-2000; 2000US-0246538.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246540.
 PR 08-NOV-2000; 2000US-0246541.
 PR 08-NOV-2000; 2000US-0246542.
 PR 08-NOV-2000; 2000US-0246543.
 PR 08-NOV-2000; 2000US-0246544.
 PR 08-NOV-2000; 2000US-0246545.
 PR 08-NOV-2000; 2000US-0246546.
 PR 08-NOV-2000; 2000US-0246547.
 PR 08-NOV-2000; 2000US-0246548.
 PR 08-NOV-2000; 2000US-0246549.
 PR 08-NOV-2000; 2000US-0246550.
 PR 08-NOV-2000; 2000US-0246551.
 PR 08-NOV-2000; 2000US-0246552.
 PR 08-NOV-2000; 2000US-0246553.
 PR 08-NOV-2000; 2000US-0246554.
 PR 08-NOV-2000; 2000US-0246555.
 PR 08-NOV-2000; 2000US-0246556.
 PR 08-NOV-2000; 2000US-0246557.
 PR 08-NOV-2000; 2000US-0246558.
 PR 08-NOV-2000; 2000US-0246559.
 PR 08-NOV-2000; 2000US-0246560.
 PR 08-NOV-2000; 2000US-0246561.
 PR 08-NOV-2000; 2000US-0246562.
 PR 08-NOV-2000; 2000US-0246563.
 PR 08-NOV-2000; 2000US-0246564.
 PR 08-NOV-2000; 2000US-0246565.
 PR 08-NOV-2000; 2000US-0246566.
 PR 08-NOV-2000; 2000US-0246567.
 PR 08-NOV-2000; 2000US-0246568.
 PR 08-NOV-2000; 2000US-0246569.
 PR 08-NOV-2000; 2000US-0246570.
 PR 08-NOV-2000; 2000US-0246571.
 PR 08-NOV-2000; 2000US-0246572.
 PR 08-NOV-2000; 2000US-0246573.
 PR 08-NOV-2000; 2000US-0246574.
 PR 08-NOV-2000; 2000US-0246575.
 PR 08-NOV-2000; 2000US-0246576.
 PR 08-NOV-2000; 2000US-0246577.
 PR 08-NOV-2000; 2000US-0246578.
 PR 08-NOV-2000; 2000US-0246579.
 PR 08-NOV-2000; 2000US-0246580.
 PR 08-NOV-2000; 2000US-0246581.
 PR 08-NOV-2000; 2000US-0246582.
 PR 08-NOV-2000; 2000US-0246583.
 PR 08-NOV-2000; 2000US-0246584.
 PR 08-NOV-2000; 2000US-0246585.
 PR 08-NOV-2000; 2000US-0246586.
 PR 08-NOV-2000; 2000US-0246587.
 PR 08-NOV-2000; 2000US-0246588.
 PR 08-NOV-2000; 2000US-0246589.
 PR 08-NOV-2000; 2000US-0246590.
 PR 08-NOV-2000; 2000US-0246591.
 PR 08-NOV-2000; 2000US-0246592.
 PR 08-NOV-2000; 2000US-0246593.
 PR 08-NOV-2000; 2000US-0246594.
 PR 08-NOV-2000; 2000US-0246595.
 PR 08-NOV-2000; 2000US-0246596.
 PR 08-NOV-2000; 2000US-0246597.
 PR 08-NOV-2000; 2000US-0246598.
 PR 08-NOV-2000; 2000US-0246599.
 PR 08-NOV-2000; 2000US-0246600.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-465566/50.

P-PSDB: AAU23090.

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases

Claim 4; SEQ ID No 186; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis).

CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polymorphisms of the invention can also be used in gene therapy.
 CC AA540785-AAA541684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences

XX Sequence 1967 BP: 723 A; 344 C; 416 G; 476 T; 8 other:

Query Match 55.8%; Score 1878; DB 22; Length 1967;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1941; Conservative 5; Mismatches 18; Indels 7; Gaps 5;

Oy 1047 AGAGATGTGGCAGCAGAGCATCCCGAGCCAGACCTCCAGCTCAGGCTTACCAAT 1106
 Db 2 AGAGATGTGGCAGCAGAGCATCCCGAGCCAGACCTCCAGCTCAGGCTTACCAAT 61
 Oy 1107 GGAAGTGGCCAGCGCTTGGAGGAGAAATATCATCTGCTGCTTACAGGGAG 1166
 Db 62 GGAAGTGGCCAGCGCTTGGAGGAGAAATATCATCTGCTGCTTACAGGGAG 121
 Oy 1167 TGGAAAAACAGAGTGGCTGTTTACATTGCCAGAGTCACTTAGACAGAGAAAAAGC 1226
 Db 122 TGGAAAAACAGAGTGGCTGTTTACATTGCCAGAGTCACTTAGACAGAGAAAAAGC 181
 Oy 1227 ATCTAGCCTGGAAAAATTATAGTTCTTGTCAATAGTACTGCTAGTGAACAGCTCTT 1286
 Db 182 ATCTAGCCTGGAAAAATTATAGTTCTTGTCAATAGTACTGCTAGTGAACAGCTCTT 241
 Oy 1287 CCGAAGAGTTCACACCATTTTGAAGAAATGATGCTGTTATGATTAAGTGGTGA 1346
 Db 242 CCGAAGAGTTCACACCATTTTGAAGAAATGATGCTGTTATGATTAAGTGGTGA 301
 Oy 1347 TACCACTGAAAAATATATATTCAGAAAGTTGTCAGAGTCTGATATTAATCATAC 1406
 Db 302 TACCACTGAAAAATATATATTCAGAAAGTTGTCAGAGTCTGATATTAATCATAC 361
 Oy 1407 AGCTAAATCCTTGAAGAACTCCCTCTTAACCTTGAAGAAATGAGAGAGATCTGCTTCA 1466
 Db 362 AGCTAAATCCTTGAAGAACTCCCTCTTAACCTTGAAGAAATGAGAGAGATCTGCTTCA 421
 Oy 1467 ATTGTACAGCTTTCCTCATATATCATGATGATGATGATGATGATGATGATGATGAT 1526
 Db 422 ATTGTACAGCTTTCCTCATATATATCATGATGATGATGATGATGATGATGATGATGAT 481
 Oy 1527 GTATATAATCATGAGGCAATTTATGATGAGAGAGTGAAGAAATAGAGTCAAGAA 1586
 Db 482 GTATATAATCATGAGGCAATTTATGATGAGAGAGTGAAGAAATAGAGTCAAGAA 541
 Oy 1587 AGAAAAACAACAGAGATTCCTCCCTCTCAGATCTGAGACTACAGCTTCACTGGTGT 1646
 Db 542 AGAAAAACAACAGAGATTCCTCCCTCTCAGATCTGAGACTACAGCTTCACTGGTGT 601
 Oy 1647 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAAGACATTTAAAGCTATGAGCAATCT 1706
 Db 602 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAAGACATTTAAAGCTATGAGCAATCT 661
 Oy 1707 TGATGATTTACTATTAAACTGTTAAAGAAAACTTATCACTGAGTGAAGAAACAAATACA 1766
 Db 662 TGATGATTTACTATTAAACTGTTAAAGAAAACTTATCACTGAGTGAAGAAACAAATACA 721
 Oy 1767 GAGAGCATGAGAAAGTTGGCATTTGCAGATGCAACCGAGAGATTCATTTAAAGAGAA 1826
 Db 722 GAGAGCATGAGAAAGTTGGCATTTGCAGATGCAACCGAGAGATTCATTTAAAGAGAA 781
 Oy 1827 ACTTCTAGAAATATGACAGAGATTCAAACTATGTCAAATGATGATGCAATTCGATTT 1886
 Db 782 ACTTCTAGAAATATGACAGAGATTCAAACTATGTCAAATGATGATGCAATTCGATTT 841
 Oy 1887 TGGAACTCAACCTATGAGCAATGGGCAATTCAAATGAAAAAAGCTGAAAAAAGG 1946
 Db 1946 TGGAACTCAACCTATGAGCAATGGGCAATTCAAATGAAAAAAGCTGAAAAAAGG 1946

Db 842 TGAAGTCAACCCATATGAAACATGGGCCATTCAATGGAAGAAAAAGCTGCAGAAAGAG 901
 Oy 1947 AAATCGCAAGAGACGTGTTGTGCAAGACATTTGAGAGATGATAGAGCTTACAAAT 2006
 Db 902 AAATCGCAAGAGACGTGTTGTGCAAGACATTTGAGAGATGATAGAGCTTACAAAT 961
 Oy 2007 TAATGACCAATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2066
 Db 962 TAATGACCAATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021
 Oy 2067 GAAAGATGAGAAAGTTGGAGTCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 2126
 Db 1022 GAAAGATGAGAAAGTTGGAGTCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1081
 Oy 2127 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2186
 Db 1082 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1141
 Oy 2187 TAGATTTCTCATGCTTATTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 2246
 Db 1142 TAGATTTCTCATGCTTATTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1201
 Oy 2247 CCGAAGATGAGAAATGAGAAAGTGCAGCAATTAAGAAATACATTAATGAGTAAATATAC 2306
 Db 1202 CCGAAGATGAGAAATGAGAAAGTGCAGCAATTAAGAAATACATTAATGAGTAAATATAC 1261
 Oy 2307 TAGAGCTGAGAAATGAGAGAGAGAGATTTTACAAAAACAGAGAGATGATATGC 2366
 Db 1262 TAGAGCTGAGAAATGAGAGAGAGAGATTTTACAAAAACAGAGAGATGATATGC 1321
 Oy 2367 GCTTCCGAGAGATTTACTGAAATGAGAAATTTGCTGAAGTACAGTCAAGTCAACA 2426
 Db 1322 GCTTCCGAGAGATTTACTGAAATGAGAAATTTGCTGAAGTACAGTCAAGTCAACA 1381
 Oy 2427 TGTGATGAGAGCTGAGACACAGAGTGAAGTCAAAACCATGACAGAGATGAGAAATGAGAA 2486
 Db 1382 TGTGATGAGAGCTGAGACACAGAGTGAAGTCAAAACCATGACAGAGATGAGAAATGAGAA 1441
 Oy 2487 AGTCAATGTAATTTTGGCACTGAGAAAAATCAATCTGTTATGCTTACCAATGAGAGAA 2546
 Db 1442 AGTCAATGTAATTTTGGCACTGAGAAAAATCAATCTGTTATGCTTACCAATGAGAGAA 1501
 Oy 2547 AGAAGTGTGATATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2606
 Db 1501 AGAAGTGTGATATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 Oy 2607 AATAGCCATGGTCCAGGCGCTGCTGAGCCAGAGCTGATGAGAGACATGATGCTGT 2666
 Db 1561 AATA-CCATGGTCCA-GCCCGTGGTGGAGTGAAGCTGATGAGAGACATGATGCTGT 1617
 Oy 2667 TGTCTACAGTGGTTCAGAGATTTATGCAATGAGAGAGTAAATGATTTCCGAGAA 2726
 Db 1618 TGTCTACAGTGGTTCAGAGATTTATGCAATGAGAGAGTAAATGATTTCCGAGAA 1677
 Oy 2727 GATGATTAAGCTATACATTTGTTCAAAATATGAAACAGAGAGATGATGATGATGATGAT 2786
 Db 1678 GATGATTAAGCTATACATTTGTTCAAAATATGAAACAGAGAGATGATGATGATGATGAT 1744
 Oy 2787 TTTGGAATTAAGATGAGAAAGTATTAAGAAAGAAATGAAACCAAGAGAAATATTGC 2846
 Db 1735 TTTGGAATTAAGATGAGAAAGTATTAAGAAAGAAATGAAACCAAGAGAAATATTGC 1794
 Oy 2847 CAAGATTAAGAAATTAACCATCACTAATACTTTCTTTGCAAAAACTGAGAGAGT 2906
 Db 1795 CAAGATTAAGAAATTAACCATCACTAATACTTTCTTTGCAAAAACTGAGAGAGT 1854
 Oy 2907 AGCTGTTCTGGGAGAGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2966
 Db 1855 AGCTGTTCTGGGAGAGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914
 Oy 2967 AGAATTCAGAGAACTTACATTTGTAAGAGAAACAGAGCACTGCAAGAGAA 3027
 Db 1915 AGAATTCAGAGAACTTACATTTGTAAGAGAAACAGAGCACTGCAAGAGAA 1975

RESULT 5
AAI59285

| ID | AAI59285 | standard; | CDNA; | 1557 | BP. |
|----|----------|-----------|-------|------|-----|
| YY | | | | | |

AC AA159285;
YY

DT 22-OCT-2001 (first entry)
 VY

Human polynucleotide SEQ ID NO 1488.

KM Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer
KM peripheral nervous system; neotrophic; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; cholelithiasis;
chemokinesis; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.

OS Homo sapiens.

PN W0200153312-A1.

PD 26-JUL-2001

PF 26-DEC-2000; 2000WO-US34263
VY

PR 21-JAN-2000; 2000US-0488725
PR 25-APR-2000; 2000US-0553317

PR 09-JUL-2000; 2000US-0598042
PR 10-JUL-2000; 2000US-0620313

PR 03-AUG-2000; 2000US-0653450
PR 14-SEP-2000; 2000US-0653101

PR 19-OCT-2000; 2000US-0693036
PR 29-NOV-2000; 2000US-0737344

XX
XX
DA (HYGE-) HYSED TMC

| | | | | | |
|----|------|----|-------|-------|-----|
| XX | Tana | VT | 140 C | Asund | 1 V |
| PI | | | | | |

PI Wang J, Wang Z, Wehrman T
PI Zhang OA, Zhou P, Goodrich

XX
DB WPT: 2001-442253/47

DR P-PSDB; AAM40129.
XX

PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

XX
PS
claim 1: SEO ID NO

CC The invention relates

the encoded polypeptides (AAM38642-AAM42213) with neurotropic, immunosuppressant and cytostatic activity. The polynucleotides are

of the invention may be used to treat disease in gene therapy. A composition containing

CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system disorders, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammatory, leukemias and
CC C.N.S disorders.
CC Note: the sequence data for this patent did not form part of the printed
CC specification.

SQ Sequence 1557 BP; 581 A; 248 C; 320 G; 408 T; 0 other

| Query Match | 43.48; | Score 1460.2; | DB 22; | Length 1557; |
|-------------|--------|---------------|--------|--------------|
| 80.00% | 80.00% | 80.00% | 80.00% | 80.00% |

Matches 1476; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

| | | | |
|----|------|--|------|
| QY | 1882 | GATTTGGAGACTCAACCTTATGAACATGGGGCACTTCAATGGAAAAAAATTTTCCAAA | 1944 |
| DB | 61 | CATTTGGAACTCAACCTTATGAACATGGGGCACTTCAATGGAAAAAAATTTTCCAAA | 120 |
| QY | 1942 | AAAGAAATGCCAAGAACGCTGTTTGTGCAGAACATTTGAGAGAGTACATATGACCTTA | 2001 |
| DB | 121 | GAGGAAATTCGGGAGAACGCTGTTTGTGCAGAACATTTGAGAGAGTACATATGACCTTA | 180 |
| QY | 2002 | CAAAATTAATGACAAATTTGCAATGATGATGGCGATACATCATCTTTGAAACTCTTATAT | 206 |
| DB | 181 | CAAAATTAATGACAAATTTGCAATGATGATGGCGATACATCATCTTTGAAACTCTTATAT | 240 |
| QY | 2062 | GAGAGAAAGATTAAGAGTTTGCAGTCAATGAGAGATGATAGTGATGAGGCTATATGAT | 2121 |
| DB | 241 | GAGAGAAAGATTAAGAGTTTGCAGTCAATGAGAGATGATAGTGATGAGGCTATATGAT | 300 |
| QY | 2122 | GAGTATTGCTGATGGTGATGAAGATGAGAGATGATTAAAGAACTTTGAAATTTGTGAA | 2181 |
| DB | 301 | GAGTATTGCTGATGGTGATGAAGATGAGAGATGATTAAAGAACTTTGAAATTTGTGAA | 360 |
| QY | 2182 | ACACATA-GATTTCATGACTTATTATTTTGGAAACCAATAAATTTGAAAGATGTGC | 2244 |
| DB | 361 | ACAGATACGATTTCATGACTTATTATTTTGGAAACCAATAAATTTGAAAGATGTGC | 420 |
| QY | 2241 | TGAAAACCCAGATATGAAAAATGAAAAAGCTGACCCAAATTTAGAAATACCATATGACGA | 2300 |
| DB | 421 | TGAAAACCCAGATATGAAAAATGAAAAAGCTGACCCAAATTTAGAAATACCATATGACGA | 480 |
| QY | 2301 | ATATACTGAGACTGAGGATACAGACAGAGAAATATCTTTACAAAAACCGATATAGTGC | 2360 |
| DB | 481 | ATATACTGAGACTGAGGATACAGACAGAGAAATATCTTTACAAAAACCGATATAGTGC | 540 |
| QY | 2361 | ATATGCGCTTCCAGTGATTAAGTAAATGAAAAATTTGCTGAGTAGCAATTAAGGC | 2420 |
| DB | 541 | ATATGCGCTTCCAGTGATTAAGTAAATGAAAAATTTGCTGAGTAGCAATTAAGGC | 600 |
| QY | 2421 | CCACCATCTGATTTGAGACTGGACACACGCACTGATGTTCAAAACCATGACACATAGAAC | 2480 |
| DB | 601 | CCACCATCTGATTTGAGACTGGACACACGCACTGATGTTCAAAACCATGACACATAGAAC | 660 |
| QY | 2481 | AAAAGAACTCTTATGTAATTTTGGCACTGAAAAATCAATCTGCTTATGCTACACAT | 2540 |
| DB | 661 | AAAAGAACTCTTATGTAATTTTGGCACTGAAAAATCAATCTGCTTATGCTACACAT | 720 |
| QY | 2541 | GGCAGAAAGACTCTGATATTAAGAAGTAACTGATTCCTGATAGGATATGATGAC | 2600 |
| DB | 721 | GGCAGAAAGACTCTGATATTAAGAAGTAACTGATTCCTGATAGGATATGATGAC | 780 |
| QY | 2601 | CAATGAAATATACCATGCTGCCAGGCCGTGTCGAGCCAGACAGCTGATGAGAGATTAAGCT | 2660 |
| DB | 781 | CAATGAAATATACCATGCTGCCAGGCCGTGTCGAGCCAGACAGCTGATGAGAGATTAAGCT | 840 |
| QY | 2661 | CCTGGTCTCTACAGTGGTTCAGAGATTAATGCAACATGAGACAGTTAATGATTTGAGGA | 2720 |
| DB | 841 | CCTGGTCTCTACAGTGGTTCAGAGATTAATGCAACATGAGACAGTTAATGATTTGAGGA | 900 |
| QY | 2721 | GAAAGATGATTAAGACTATACATGTTGTTCAAAATATGAAAAACGAGAGATATGCTCA | 2780 |
| DB | 901 | GAAAGATGATTAAGACTATACATGTTGTTCAAAATATGAAAAACGAGAGATATGCTCA | 960 |
| QY | 2781 | TAAGATTTTGGATTTACAGATGCAAACTATATGAAAAAAGAAAAATGAAACATTAAGAAA | 2840 |
| DB | 961 | TAAGATTTTGGATTTACAGATGCAAACTATATGAAAAAAGAAAAATGAAACATTAAGAAA | 1020 |
| QY | 2841 | TATGGCCAAAGATTAAGAAATTAACCATCACTAATAACTTCCTTTGAAATAAATGAG | 2900 |
| DB | 1021 | TATGGCCAAAGATTAAGAAATTAACCATCACTAATAACTTCCTTTGAAATAAATGAG | 1080 |
| QY | 2901 | TGTGCTACGCCCTGTTCTGGGAGACATATCCATGTAAATTTGAGAAAAATGACATCACTAATAT | 2960 |
| DB | 1081 | TGTGCTACGCCCTGTTCTGGGAGACATATCCATGTAAATTTGAGAAAAATGACATCACTAATAT | 1140 |
| QY | 2961 | GACCCAGAAATTCAGGAACCTTACTATGTTAAGACAAAAAAGCACTGCAAAATTAAGTG | 3020 |

Db 1141 GACCCGAGATTCAGGAGCTTACATGTAGAGAAACAAAGACTGTAAGAGAGTG 1200
 QY 3021 TCGCGACTATCAATTAATAGTGAATCATCTGCAAAATGTCGCGAGGCTTGGGAAACAT 3080
 Db 1201 TCGCGACTATCAATTAATAGTGAATCATCTGCAAAATGTCGCGAGGCTTGGGAAACAT 1260
 QY 3081 GATGTGCAAAAGGCTTATGATTTGCTGCTCAAAATAGCAATTTTATGTCGTTT 3140
 Db 1261 GATGTGCAAAAGGCTTATGATTTGCTGCTCAAAATAGCAATTTTATGTCGTTT 1320
 QY 3141 CAAAAATTAATCAACAAAGAAACATATCAAAAAAGTGGTGAATTAACCTTACATTTTC 3200
 Db 1321 CAAAAATTAATCAACAAAGAAACATATCAAAAAAGTGGTGAATTAACCTTACATTTTC 1380
 QY 3201 CAATCTGACTATGCAATGCTGTTTATGATGATGAGATTAACATGATGAGAA 3260
 Db 1381 CAATCTGACTATGCAATGCTGTTTATGATGATGAGATTAACATGATGAGAA 1440
 QY 3261 TTTCTTTAAATTAATCAATGATTAACATTTATGATGATGATGATGATGATG 3320
 Db 1441 TTTCTTTAAATTAATCAATGATTAACATTTATGATGATGATGATGATGATG 1500
 QY 3321 CTACAGACTGACATTAAGATCAATTAATGATTTTACTCTG 3385
 Db 1501 CTACAGACTGACATTAAGATCAATTAATGATTTTACTCTG 1541
 RESULT 6
 ABA04916
 ID ABA04916 standard; DNA; 1443 BP.
 XX ABA04916;
 XX
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH16 related DNA sequence #3.
 XX
 KW Human: RH16; RNA helicase; cytosolic; vironcide; anti-HIV;
 immunosuppressive; immunostimulatory; antileukemic; a triarthritic;
 antileukosclerotic; osteopathic; antidiabetic; hepatotropic;
 antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;
 autoimmune disease; graft rejection; vaccine; ds.
 KW
 OS Homo sapiens.
 XX
 PN WO200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 11-MAY-2001; 2001MO-FR01441.
 XX
 PR 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 XX
 PI Bahr G, Cocude C, Capron A;
 XX
 DR WPI; 2002-082898/11.
 PT
 PT New polypeptide, useful for treating and diagnosing cancer or
 PT inflammation, and drug screening, comprises a human polynucleotide
 PT homologous to RNA helicase
 XX
 PS Example 2; Page 94-95; 114pp; French.
 XX
 CC The present invention relates to human RH16 (see AAM4778). RH16 is a
 CC 116kDa protein and has homology to RNA helicases (DEXH box). RH16 and
 CC its coding sequence are useful for treating cancer; acute or chronic
 CC infections (especially by HIV or hepatitis B or C); inherited genetic
 CC diseases; (auto)immune diseases (particularly rheumatoid arthritis,
 CC arteriosclerosis, osteoporosis and diabetes, but many others listed) and

CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.
 XX

SO Sequence 1443 BP; 508 A; 261 C; 313 G; 360 T; 1 other:

Query Match 42.7%; Score 1435.6; DB 24; Length 1443;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 687 GAAAGAAAAGTGTCTCTGCAATTTGCAATGTTCTTGTGTAAGAGAAATGAAT 746
 Db 1 GAAAGAAAAGTGTCTCTGCAATTTGCAATGTTCTTGTGTAAGAGAAATGAAT 60
 QY 747 TGTCCAGAGTAACAGGCTGATTTGCTGAAAGCAATGAGATTTGAAATTTTC 806
 Db 61 TGTCCAGAGTAACAGGCTGATTTGCTGAAAGCAATGAGATTTGAAATTTTC 120
 QY 807 ACAGATGATGCTCTCAAGTGAAGAGCAACTTCTTCAACCAAGTTCAGTCAATCT 866
 Db 121 ACAGATGATGCTCTCAAGTGAAGAGCAACTTCTTCAACCAAGTTCAGTCAATCT 180
 QY 867 GGAGAGAGAGTCTGGGCAATGAGATTAATCAATCAATCAATCTTTTGCAGATCTC 926
 Db 181 GGAGAGAGAGTCTGGGCAATGAGATTAATCAATCAATCAATCTTTTGCAGATCTC 240
 QY 927 TGTAGTTTCAAGATGAGACACAAAGTTTGGAGAGAGTGTGACCTGTTAGAGAG 986
 Db 241 TGTAGTTTCAAGATGAGACACAAAGTTTGGAGAGAGTGTGACCTGTTAGAGAG 300
 QY 987 TCTTGACATTAACAGCAATGAGGAGTATTCAGGACCAATGGAGATGATTCAGAT 1046
 Db 301 TCTTGACATTAACAGCAATGAGGAGTATTCAGGACCAATGGAGATGATTCAGAT 360
 QY 1047 ACAGATGAGGAGCAAGAGCATTCCTGGAGGACCAATTCAGGACCAATTCAGAT 1106
 Db 361 ACAGATGAGGAGCAAGAGCATTCCTGGAGGACCAATTCAGGACCAATTCAGAT 420
 QY 1107 GGAAGTGTGCGCAGCAGCTTGGAGGAGAAATTCATCATCTGCTCCCTTACAGGAG 1166
 Db 421 GGAAGTGTGCGCAGCAGCTTGGAGGAGAAATTCATCATCTGCTCCCTTACAGGAG 480
 QY 1167 TGGAAAAACAGAGTGGCTGTTTACATTTGCCAAGATCACTTAGCAAGAAACAAAAG 1226
 Db 481 TGGAAAAACAGAGTGGCTGTTTACATTTGCCAAGATCACTTAGCAAGAAACAAAAG 540
 QY 1227 ATCTGAGCTGGAAGAAATTAATGTTCTGTCAATTAAGTACTGCTAGTTGAATAC 1286
 Db 541 ATCTGAGCTGGAAGAAATTAATGTTCTGTCAATTAAGTACTGCTAGTTGAATAC 600
 QY 1287 CCGCAAGAGTTCGAAGCAATTTTGAAGAAATGATGATGATTTGATTAAGAGAG 1346
 Db 601 CCGCAAGAGTTCGAAGCAATTTTGAAGAAATGATGATGATTTGATTAAGAGAG 660
 QY 1347 TACCACTGAAGAAATTAATTCAGAAAGTGTCAAGTCTGTATTAATTAATAGAC 1406
 Db 661 TACCACTGAAGAAATTAATTCAGAAAGTGTGTCAAGTCTGTATTAATTAATAGAC 720
 QY 1407 AGCTCAATCTTGAAGAACTCCCTCTTAACCTTGAAGAAATGAGAGAAATGCTGAT 1466
 Db 721 AGCTCAATCTTGAAGAACTCCCTCTTAACCTTGAAGAAATGAGAGAAATGCTGAT 780
 QY 1467 ATTGTGAGCTTTTCCCTCATTAATGATGATGATGATGATGATGATGATGATGAT 1526
 Db 781 ATTGTGAGCTTTTCCCTCATTAATGATGATGATGATGATGATGATGATGATGAT 840
 QY 1527 GTATATAACATCAAGAGCATTAATTTGATGAGAGAGTGAAGAAACATAGAT 1586
 Db 841 GTATATAACATCAAGAGCATTAATTTGATGAGAGAGTGAAGAAACATAGAT 900
 QY 1587 AGAAAAAGAGAGATTTCCCTCTGATGATGATGATGATGATGATGATGATGATGAT 1646
 Db 901 AGAAAAAGAGAGATTTCCCTCTGATGATGATGATGATGATGATGATGATGATGAT 960

| QY | 1647 | TGGAGGGGGCCACGAGACGACGAAAGCTGAAAGAACATTTTAAAGCTATGGCAATCT | 1760 |
|----------|---|---|------|
| QY | 1647 | TGGAGGGGGCCACGAGACGACGAAAGCTGAAAGAACATTTTAAAGCTATGGCAATCT <td>1760</td> | 1760 |
| Db | 961 | TGGAGGGGGCCACGAGACGACGAAAGCTGAAAGAACATTTTAAAGCTATGGCAATCT | 1020 |
| QY | 1707 | TGATGACATTTACTATTAAACGTGTTAAAGAAAAACCTTGAATCACTGAAAAATCAATACA | 1766 |
| Db | 1021 | TGATGACATTTACTATTAAACGTGTTAAAGAAAAACCTTGAATCACTGAAAAATCAATACA | 1080 |
| QY | 1767 | GGAGCCATGCAGAAAGTTTGGCCATTGCGAGTGCACCCAGAGAGATTCATTTAAAGAGAA | 1826 |
| Db | 1081 | GGAGCCATGCAGAAAGTTTGGCCATTGCGAGTGCACCCAGAGAGATTCATTTAAAGAGAA | 1140 |
| QY | 1827 | ACTCTGAAATTAATGACAGAGATTTGAACTTAACTTATGTCAAATTAATGTCATGTCAGATT | 1886 |
| Db | 1141 | ACTCTGAAATTAATGACAGAGATTTGAACTTAACTTATGTCAAATTAATGTCATGTCAGATT | 1200 |
| QY | 1887 | TGGAAGTCACCCCTATGACACATGGGCCATTCCAAATGTGAAAAAAAGCTGAAAAAAAG | 1946 |
| Db | 1201 | TGGAAGTCACCCCTATGACACATGGGCCATTCCAAATGTGAAAAAAAGCTGAAAAAAAG | 1260 |
| QY | 1947 | AAATCGGAAAGAAACGTGTTTGTGCGAAGAAATTTGAGAGAGTTCATGAGACCTGCAAAAT | 2006 |
| Db | 1261 | AAATCGGAAAGAAAGTGTGTTGTGCGAAGAAATTTGAGAGAGTTCATGAGACCTGCAAAAT | 1320 |
| QY | 2007 | TAAATGACACATTCGAAATGATAGATGCGTATACATCTTGAACCTTGCATTAATGAGAA | 2066 |
| Db | 1321 | TAAATGACACATTCGAAATGATAGATGCGTATACATCTTGAACCTTGCATTAATGAGAA | 1380 |
| QY | 2067 | GAAAGATTAAGAACTTTCGACTCATTAAGAATGATAGTAGATGAGAGGTGGTAAATGATGATA | 2126 |
| Db | 1381 | GAAAGATTAAGAAAGTTTGCAGCTCATTAAGAATGATAGTAGATGAGAGGTGGTAAATGATGATA | 1440 |
| QY | 2127 | TTG 2129 | |
| Db | 1441 | TTG 1443 | |
| RESULT 7 | | | |
| AA161071 | AA161071 standard; cDNA: 1382 BP. | | |
| XX | AA161071: | | |
| AC | 22-OCT-2001 (first entry) | | |
| XX | Human polynucleotide SEQ ID NO 5060. | | |
| DE | Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer; | | |
| KW | peripheral nervous system; neuropathy; central nervous system; CNS; | | |
| KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; | | |
| KW | amyotrophic lateral sclerosis; Shy-Drager syndrome; ch-motocytic; | | |
| KW | hemophilic; thrombolytic; drug screening; arthritis; inflammation; | | |
| KW | leukemia; ss. | | |
| OS | Homo sapiens. | | |
| XX | MO20015312-Al. | | |
| PN | 26-JUL-2001. | | |
| XX | 26-DEC-2000; 2000MO-US34263. | | |
| XX | 21-JAN-2000; 2000US-0488725. | | |
| PR | 25-APR-2000; 2000US-0552317. | | |
| PR | 09-JUL-2000; 2000US-0598042. | | |
| PR | 19-JUL-2000; 2000US-0620312. | | |
| PR | 03-AUG-2000; 2000US-0653450. | | |
| PR | 14-SEP-2000; 2000US-0662191. | | |
| PR | 19-OCT-2000; 2000US-0693036. | | |
| PR | 29-NOV-2000; 2000US-0727344. | | |
| XX | (HYSE-) HYSEQ INC. | | |

[illegible]

| | | | | | | |
|--|--------------------------|---|------|--|--|--|
| OY | 2643 | TGATGAGAGCGACCTACGCTCGTGGTGTCTGACAGTGGTTCAGAGAGTAACTAACATGATGAC | 2702 | | | |
| Db | 541 | TGATGAGAGCGACCTACGCTCGTGGTGTCTGACAGTGGTTCAGAGAGTAACTAACATGATGAC | 600 | | | |
| OY | 2703 | AGTTAATGATATTTCCGAGAGAAAGATGATGATATAAAGCTATACATTGTGTTTAAATATGAA | 2762 | | | |
| Db | 601 | AGTTAATGATATTTCCGAGAGAAAGATGATGATATAAAGCTATACATTGTGTTTAAATATGAA | 660 | | | |
| OY | 2763 | ACCGAGAGAGTATGCTCTATTAAGATTTTGGAAATTTACAGATGCAAAATATATATGAAAAATA | 2822 | | | |
| Db | 661 | ACCGAGAGAGTATGCTCTATTAAGATTTTGGAAATTTACAGATGCAAAATATATATGAAAAATA | 720 | | | |
| OY | 2823 | AATGAAAAACCAAGAGAAATATTGCGCAAGCATTAACAAGATTAACCCATCACTTAATAACATT | 2882 | | | |
| Db | 721 | AATGAAAAACCAAGAGAAATATTGCGCAAGCATTAACAAGATTAACCCATCACTTAATAACATT | 780 | | | |
| OY | 2883 | CCTTTGGCAAAAATGCGAGTGTGCTAGCCGTGTTCTGGGGAAGATATGCACTTAATGATAA | 2942 | | | |
| Db | 781 | CCTTTGGCAAAAATGCGAGTGTGCTAGCCGTGTTCTGGGGAAGATATGCACTTAATGATAA | 840 | | | |
| OY | 2943 | AATGATATGAGTCAATATGAGCCGCAAGATTAACAAGAACTTTCATCTGTAAACAGAAAAATA | 3002 | | | |
| Db | 841 | AATGATATGAGTCAATATGAGCCGCAAGATTAACAAGAACTTTCATCTGTAAACAGAAAAATA | 900 | | | |
| OY | 3003 | AGCACTGCAAAAGAAAGTGTCCGCACTATGCAAAATTAATGCTGAATCATCTGCAAAATGTGG | 3062 | | | |
| Db | 901 | AMCACTGCAAAAGAAAGTGTCCGCACTATGCAAAATTAATGCTGAATCATCTGCAAAATGTGG | 960 | | | |
| OY | 3063 | CCAGGCTTGGGGAGCAATGATGTGTGCACAAAGGCTTAGATTGGCTTGCTGCAAAATAG | 3122 | | | |
| Db | 961 | CCAGGCTTGGGGAGCAATGATGTGTGCACAAAGGCTTAGATTGGCTTGCTGCAAAATAG | 1020 | | | |
| OY | 3123 | GAATTTGTGATGTGTTTCCAAAATTAATTCACAAAGAAACATACAAAGAAAGTGGGTGA | 3182 | | | |
| Db | 1021 | GAATTTGTGATGTGTTTCCAAAATTAATTCACAAAGAAACATACAAAGAAAGTGGGTGA | 1080 | | | |
| OY | 3183 | ATTACCTATACATATTTCCCAATCTTGACATTCAGATTCGAATGCTGTTATTATTAATGATGAGA | 3242 | | | |
| Db | 1081 | ATTACCTATACATATTTCCCAATCTTGACATTCAGATTCGAATGCTGTTATTATTAATGATGAGA | 1140 | | | |
| OY | 3243 | TTACACCTGATTAAGAAATCTCTTTAAATTAACATACAGTAAACATTTAAATGATATG | 3302 | | | |
| Db | 1141 | TTACACCTGATTAAGAAATCTCTTTAAATTAACATACAGTAAACATTTAAATGATATG | 1200 | | | |
| OY | 3303 | ATTATATGATTTCATTATGATGACAGAACTGACATTAAGATCAATTAATAATGATGTTTACT | 3362 | | | |
| Db | 1201 | ATTATATGATTTCATTATGATGACAGAACTGACATTAAGATCAATTAATAATGATGTTTACT | 1260 | | | |
| OY | 3363 | CT 3364 | | | | |
| Db | 1261 | NT 1262 | | | | |
| RESULT 9 | | | | | | |
| AS41517 | standard: cDNA: 1319 BP. | | | | | |
| AS41517: | | | | | | |
| 17-DEC-2001 | (first entry) | | | | | |
| cDNA encoding novel human enzyme polypeptide #733. | | | | | | |
| Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase | | | | | | |
| lyase; hyperproliferative disorder; immunodeficiency disorder; | | | | | | |
| autoimmune disorder; neurological disorder; metabolic disorder; | | | | | | |
| inflammatory disorder; cardiovascular disorder; reproductive disorder; | | | | | | |
| blood-related disorder; infectious disorder; gene therapy; cytostatic; | | | | | | |
| anti arthritic; nephrotoxic; anticoagulant; ss. | | | | | | |
| Homo sapiens. | | | | | | |

| | | | |
|----|----|-----------------------------|--|
| XX | PN | MO200155301-A2. | |
| PD | XX | 02-AUG-2001. | |
| XX | XX | 17-JAN-2001; 2001WO-US01239 | |
| PR | XX | 31-JAN-2000; 2000US-0179065 | |
| PR | PR | 04-FEB-2000; 2000US-0180628 | |
| PR | PR | 24-FEB-2000; 2000US-0184664 | |
| PR | PR | 02-MAR-2000; 2000US-0186550 | |
| PR | PR | 16-MAR-2000; 2000US-0189674 | |
| PR | PR | 17-MAR-2000; 2000US-0190076 | |
| PR | PR | 18-APR-2000; 2000US-0198123 | |
| PR | PR | 19-MAY-2000; 2000US-0205515 | |
| PR | PR | 07-JUN-2000; 2000US-0209647 | |
| PR | PR | 28-JUN-2000; 2000US-0214866 | |
| PR | PR | 30-JUN-2000; 2000US-0216535 | |
| PR | PR | 07-JUL-2000; 2000US-0216647 | |
| PR | PR | 07-JUL-2000; 2000US-0216680 | |
| PR | PR | 11-JUL-2000; 2000US-0217487 | |
| PR | PR | 11-JUL-2000; 2000US-0217496 | |
| PR | PR | 14-AUG-2000; 2000US-0225513 | |
| PR | PR | 14-AUG-2000; 2000US-0225514 | |
| PR | PR | 14-AUG-2000; 2000US-0225566 | |
| PR | PR | 14-AUG-2000; 2000US-0225567 | |
| PR | PR | 14-AUG-2000; 2000US-0225568 | |
| PR | PR | 14-AUG-2000; 2000US-0225570 | |
| PR | PR | 14-AUG-2000; 2000US-0225547 | |
| PR | PR | 14-AUG-2000; 2000US-0225577 | |
| PR | PR | 14-AUG-2000; 2000US-0225758 | |
| PR | PR | 14-AUG-2000; 2000US-0225759 | |
| PR | PR | 18-AUG-2000; 2000US-0226279 | |
| PR | PR | 22-AUG-2000; 2000US-0226681 | |
| PR | PR | 22-AUG-2000; 2000US-0226687 | |
| PR | PR | 22-AUG-2000; 2000US-0227182 | |
| PR | PR | 30-AUG-2000; 2000US-0227009 | |
| PR | PR | 30-AUG-2000; 2000US-0228924 | |
| PR | PR | 01-SEP-2000; 2000US-0229287 | |
| PR | PR | 01-SEP-2000; 2000US-0229343 | |
| PR | PR | 01-SEP-2000; 2000US-0229345 | |
| PR | PR | 01-SEP-2000; 2000US-0229344 | |
| PR | PR | 03-SEP-2000; 2000US-0229509 | |
| PR | PR | 05-SEP-2000; 2000US-0230517 | |
| PR | PR | 06-SEP-2000; 2000US-0230433 | |
| PR | PR | 06-SEP-2000; 2000US-0230438 | |
| PR | PR | 08-SEP-2000; 2000US-0231442 | |
| PR | PR | 08-SEP-2000; 2000US-0231443 | |
| PR | PR | 08-SEP-2000; 2000US-0231444 | |
| PR | PR | 08-SEP-2000; 2000US-0231413 | |
| PR | PR | 08-SEP-2000; 2000US-0231423 | |
| PR | PR | 08-SEP-2000; 2000US-0231080 | |
| PR | PR | 08-SEP-2000; 2000US-0232081 | |
| PR | PR | 12-SEP-2000; 2000US-0231668 | |
| PR | PR | 14-SEP-2000; 2000US-0232397 | |
| PR | PR | 14-SEP-2000; 2000US-0232398 | |
| PR | PR | 14-SEP-2000; 2000US-0232399 | |
| PR | PR | 14-SEP-2000; 2000US-0232401 | |
| PR | PR | 14-SEP-2000; 2000US-0233063 | |
| PR | PR | 14-SEP-2000; 2000US-0233064 | |
| PR | PR | 14-SEP-2000; 2000US-0233065 | |
| PR | PR | 21-SEP-2000; 2000US-0234223 | |
| PR | PR | 21-SEP-2000; 2000US-0234224 | |
| PR | PR | 25-SEP-2000; 2000US-0234997 | |
| PR | PR | 25-SEP-2000; 2000US-0234998 | |
| PR | PR | 26-SEP-2000; 2000US-0234984 | |
| PR | PR | 27-SEP-2000; 2000US-0235834 | |
| PR | PR | 27-SEP-2000; 2000US-0235836 | |

| | | | |
|----|--------------------------------|----------------|--|
| PR | 29-SEP-2000 | 2000US-0236327 | |
| PR | 29-SEP-2000 | 2000US-0236367 | |
| PR | 29-SEP-2000 | 2000US-0236368 | |
| PR | 29-SEP-2000 | 2000US-0236369 | |
| PR | 02-OCT-2000 | 2000US-0236370 | |
| PR | 02-OCT-2000 | 2000US-0236371 | |
| PR | 02-OCT-2000 | 2000US-0236372 | |
| PR | 02-OCT-2000 | 2000US-0237038 | |
| PR | 02-OCT-2000 | 2000US-0237039 | |
| PR | 02-OCT-2000 | 2000US-0237040 | |
| PR | 13-OCT-2000 | 2000US-0239335 | |
| PR | 13-OCT-2000 | 2000US-0239337 | |
| PR | 20-OCT-2000 | 2000US-0240960 | |
| PR | 20-OCT-2000 | 2000US-0241121 | |
| PR | 20-OCT-2000 | 2000US-0241185 | |
| PR | 20-OCT-2000 | 2000US-0241186 | |
| PR | 20-OCT-2000 | 2000US-0241187 | |
| PR | 20-OCT-2000 | 2000US-0241808 | |
| PR | 20-OCT-2000 | 2000US-0241809 | |
| PR | 20-OCT-2000 | 2000US-0241826 | |
| PR | 01-NOV-2000 | 2000US-0244617 | |
| PR | 08-NOV-2000 | 2000US-0244617 | |
| PR | 08-NOV-2000 | 2000US-0246475 | |
| PR | 08-NOV-2000 | 2000US-0246476 | |
| PR | 08-NOV-2000 | 2000US-0246477 | |
| PR | 08-NOV-2000 | 2000US-0246478 | |
| PR | 08-NOV-2000 | 2000US-0246523 | |
| PR | 08-NOV-2000 | 2000US-0246524 | |
| PR | 08-NOV-2000 | 2000US-0246525 | |
| PR | 08-NOV-2000 | 2000US-0246526 | |
| PR | 08-NOV-2000 | 2000US-0246527 | |
| PR | 08-NOV-2000 | 2000US-0246528 | |
| PR | 08-NOV-2000 | 2000US-0246532 | |
| PR | 08-NOV-2000 | 2000US-0246609 | |
| PR | 08-NOV-2000 | 2000US-0246610 | |
| PR | 08-NOV-2000 | 2000US-0246611 | |
| PR | 08-NOV-2000 | 2000US-0246613 | |
| PR | 17-NOV-2000 | 2000US-0249207 | |
| PR | 17-NOV-2000 | 2000US-0249208 | |
| PR | 17-NOV-2000 | 2000US-0249209 | |
| PR | 17-NOV-2000 | 2000US-0249210 | |
| PR | 17-NOV-2000 | 2000US-0249211 | |
| PR | 17-NOV-2000 | 2000US-0249212 | |
| PR | 17-NOV-2000 | 2000US-0249213 | |
| PR | 17-NOV-2000 | 2000US-0249214 | |
| PR | 17-NOV-2000 | 2000US-0249215 | |
| PR | 17-NOV-2000 | 2000US-0249216 | |
| PR | 17-NOV-2000 | 2000US-0249217 | |
| PR | 17-NOV-2000 | 2000US-0249218 | |
| PR | 17-NOV-2000 | 2000US-0249244 | |
| PR | 17-NOV-2000 | 2000US-0249245 | |
| PR | 17-NOV-2000 | 2000US-0249264 | |
| PR | 17-NOV-2000 | 2000US-0249265 | |
| PR | 17-NOV-2000 | 2000US-0249297 | |
| PR | 17-NOV-2000 | 2000US-0249299 | |
| PR | 17-NOV-2000 | 2000US-0249300 | |
| PR | 01-DEC-2000 | 2000US-0250391 | |
| PR | 01-DEC-2000 | 2000US-0250392 | |
| PR | 05-DEC-2000 | 2000US-0251038 | |
| PR | 05-DEC-2000 | 2000US-0251088 | |
| PR | 05-DEC-2000 | 2000US-0251619 | |
| PR | 06-DEC-2000 | 2000US-0251479 | |
| PR | 06-DEC-2000 | 2000US-0251856 | |
| PR | 08-DEC-2000 | 2000US-0251868 | |
| PR | 08-DEC-2000 | 2000US-0251869 | |
| PR | 08-DEC-2000 | 2000US-0251889 | |
| PR | 08-DEC-2000 | 2000US-0251997 | |
| PR | 11-DEC-2000 | 2000US-0254097 | |
| PR | 05-JAN-2001 | 2001US-0259678 | |
| XX | | | |
| EA | (HUMA-) HUMAN GENOME SCI INC. | | |
| XX | | | |
| PI | Rosen CA, Barash SC, Ruben SM; | | |

XX WP1: 2001.465566/50.
DR P-PSDB: AAU23647.

XX Novel polypeptides and polynucleotides useful for diagnosis,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases.

PS Claim 4: SEQ ID NO 743; 1180pp; English.

XX The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU50785-AAU51684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC atftp.wipo.int/pub/published_pct_sequences.

XX
XX
XX Sequence 1319 BP; 491 A; 217 C; 294 G; 314 T; 3 other;

Query Match 33.6%; Score 1130.8; DB 22; Length 415;
Best Local Similarity 95.3%; Pred. No. 4.7e-270;
Matches 1253; Conservative 2; Mismatches 40; Indels 40; Gaps

QY 1493 TTGATGAATGTCATCAGACCAACAAAGAGCAGTATATATACATCATGAGGTTATT 1552
DB 24 TTGATGAATGTCATCAGACCAACAAAGAGCAGTATATATATACATCATGAGCTTATT 83

QY 1553 TGATGCAGAAAGTTGAAAAACATTAACATCAGAAAAAGAAAAACCAACACGATGATTCGCTTC 1612
DB 84 TGATGCAGAAAGTTGAAAAACATTAACATCAGAAAAAGAAAAACCAACGATGATTCGCTTC 143

QY 1613 CTCAGATACCTGGAGCTAACAGCTTCACCTGCGTGTGGAGGGCCAGAGCAAGCAAG 1672
DB 144 CTCAGATACCTGGAGCTAACAGCTTCACCTGCGTGTGGAGGGCCAGAGCAAGCAAG 203

QY 1673 CTGAAGAACACATTTTAAAAACTATGTGCCAACTTGTATGATCATTTACTATTAATAATGTTA 1732
DB 204 CTGAAGAACACATTTTAAAAACTATGTGCCAACTTGTATGATCATTTACTATTAATAATGTTA 263

QY 1733 AAGAAACCTTATCATCACTGAAAAAACCAATACAGAGCCATGCAAGAAGCTTTCGATTCG 1792
DB 264 AAGAAACCTTATCATCACTGAAAAAACCAATACAGAGCCATGCAAGAAGCTTTCGATTCG 323

QY 1793 CAGATGCACCGAGAGAAAGTGCATTTAAAGAAACTTCTAATAATATGACCAAGATTC 1852
DB 324 CAGATGCACCGAGAGAAAGTGCATTTAAAGAAACTTCTAATAATATGAAAGCAATTC 383

QY 1853 AAACCTATATGTCAAAATGAGTCCAAATGTGAGATTTTGGAACTCAACCTATGACAAAGGG 1912
DB 384 AAACCTATATGTCAAAATGAGTCCAAATGTGAGATTTTGGAACTCAACCTATGACAAAGGG 443

QY 1913 CCATTCAAATGCAAAAAAGCTGCAAAAAAAAGAAATGCAAGAGACGTGTATGTCAG 1972
DB 444 CCATTCAAATGCAAAAAAGCTGCAAAAAAAAGAAATGCAAGAGACGTGTATGTCAG 503

QY 1973 AACATTGTGGAGAGTACAAATGAGGCCCTTACAAATTAATGACCAATTTGCAATATATATG 2032
DB 504 AACATTGTGGAGAGTACAAATGAGGCCCTTACAAATTAATGACCAATTTGCAATATATATG 563

QY 2033 CGTATCTCATCTTGAACCTTCTATATGAGAGAGAAAGTATTCAGTATAG 2092
 DB 564 CGTATCTCATCTTGAACCTTCTATATGAGAGAGAAAGTATTCAGTATAG 623
 QY 2093 AGATATAGTATGAGAGGAGTATGATGATGATGATGATGATGATGATGATG 2152
 DB 624 AGATATAGTATGAGAGGAGTATGATGATGATGATGATGATGATGATGATG 683
 QY 2153 ATTTAAAGAACTTTGAAAGCTGATGAAACAGATAGATTCTCATGAAAT 2212
 DB 684 ATTTAAAGAACTTTGAAAGCTGATGAAACAGATAGATTCTCATGAAAT 743
 QY 2213 AAAACATTAATTAATGTTGAAAGAGCTGTAAGAACCAATATGAAATAGAA 2272
 DB 744 AAAACATTAATTAATGTTGAAAGAGCTGTAAGAACCAATATGAAATAGAA 803
 QY 2273 CCAATTAAGAAATACCATATGATGATGATGATGATGATGATGATGATGATG 2332
 DB 804 CCAATTAAGAAATACCATATGATGATGATGATGATGATGATGATGATGATG 863
 QY 2333 TAATCTTTACAAAACAGACAGAGTATGATGATGATGATGATGATGATGATG 2392
 DB 864 TAATCTTTACAAAACAGACAGAGTATGATGATGATGATGATGATGATGATG 923
 QY 2393 AAAAATTTGCTGAGTAGAGTCAAGCCACCATCTGATTTGAGCTGTAAG 2452
 DB 924 AAAAATTTGCTGAGTAGAGTCAAGCCACCATCTGATTTGAGCTGTAAG 983
 QY 2453 AGTTCAACCCATGACACAGATGAACAAAAGAACTCATAGTAATTTGACAG 2512
 DB 984 AGTTCAACCCATGACACAGATGAACAAAAGAACTCATAGTAATTTGACAG 1043
 QY 2513 AATTCATCTGCTTATCGTACACAGCTGACAGAGAGTATGATATTAAGATGTA 2572
 DB 1044 AATTCATCTGCTTATCGTACACAGCTGACAGAGAGTATGATATTAAGATGTA 1102
 QY 2573 ACATTTGATTCGCTTATCGTACACAGCTGACAGAGAGTATGATATTAAGATGTA 2632
 DB 1103 ACATTTGATTCGCTTATCGTACACAGCTGACAGAGAGTATGATATTAAGATGTA 1160
 QY 2633 GAGCAGAGCTGATGAGAGAGCTGATGATGATGATGATGATGATGATGATG 2692
 DB 1161 GAGCAGAGCTGATGAGAGAGCTGATGATGATGATGATGATGATGATGATG 1217
 QY 2693 ACATGAGACGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 2752
 DB 1218 -ACRTGAGACG-TATGATTTGAGAGAAATGATG-----TATTAATATCAG 1264
 QY 2753 AAAATATGAACAGAGAGATGATGATGATGATGATGATGATGATGATGATG 2807
 DB 1265 GTCAAAATATTAACAGAGAGATGATGATGATGATGATGATGATGATGATG 1318
 RESULT 10
 AAS01149
 ID AAS01149 standard; DNA; 1258 BP.
 XX
 AC AAS01149;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Interferon induced nucleic acid, IFN4.
 XX
 KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KW haematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN4; ds.
 OS Homo sapiens.
 XX

FH Key Location/Qualifiers
 FT CDS 1..915
 FT /tag= a
 FT /product= "IFN4"
 FT /note= "Interferon induced polypeptide"
 PN MO200118208-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000WO-US24704.
 XX
 PR 08-SEP-1999; 98US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 XX
 PA (CURA-) CURAGEN CORP.
 PA (BIOJ) BIOGEN INC.
 XX
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR P-PSDB: AAU00296.
 PT New interferon induced polypeptides and polynucleotides, useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease.
 PS Claim 9; Page 29-32; 134pp; English.
 XX
 CC The sequence represents Interferon induced nucleic acid, IFN4. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating neurological
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SO Sequence 1258 BP; 399 A; 267 C; 276 G; 316 T; 0 other:
 Query Match 23.9%; Score 803; DB 22; Length 1258;
 Best Local Similarity 95.5%; Pred. No. 9,2e-189;
 Matches 853; Conservative 0; Mismatches 0; Indels 40; Gaps 1;
 QY 45 GCACAGACGGTAGACCTGCTTCTTAAGTGGGACGGACAGCGACG/AA:ATTTC 104
 DB 55 GCACAGACGGTAGACCTGCTTCTTAAGTGGGACGGACAGCGACG/AA:ATTTC 114
 QY 105 CCGTCCCGGACAGCAAGCAACATTCCTTGGGAGAAACCTTCCTCTTCAAGAAAG 164
 DB 115 CCGTCCCGGACAGCAAGCAACATTCCTTGGGAGAAACCTTCCTCTTCAAGAAAG 174
 QY 165 AAAGATGTGAATGGTATTCACAGAGAGAAATTTCCGATATCATCTCTGAG 224
 DB 175 AAAGATGTGAATGGTATTCACAGAGAGAAATTTCCGATATCATCTCTGAG 234

285 TGCAGAGTGAAGAGAGATTCAGAGAGACAGTGGCCAGCTCCGGGAACTCCAGGCACT 344
 |||||||
 Db 269 --CAGAGGTAAAGAGAGAGATTCAGAGAGACAGTGGCCAGCTCCGGGAACTCCAGGCACT 326
 345 TGAAGTCTGCTGCTGAGCAGCTTGGAGAGAGGAGTCTGGACCTTGGTGGTCTGGGAAAT 404
 |||||||
 Db 327 TGAAGTCTGCTGCTGAGCAGCTTGGAGAGAGGAGTCTGGACCTTGGTGGTCTGGGAAAT 386
 405 CGTGGAGGCGCTCCGGAGAACCGGACCGCTCTGGCCCGCCGCTACATCTTCCCTGAGCT 464
 |||||||
 Db 387 CGTGGAGGCGCTCCGGAGAACCGGACCGCTCTGGCCCGCCGCTACATCTTCCCTGAGCT 416
 465 CACGAGCTTCCCTCCAGCTGTTTGAAGAACGCTCATGATGATATCTCACTGCTGAA 524
 |||||||
 Db 447 CACGAGCTTCCCTCCAGCTGTTTGAAGAACGCTCATGATGATATCTCACTGCTGAA 506
 525 CCTCCTTCCAGCCCTCTGCTGGAGCAAGCTTCTAGTATGAGAGCTCTTCTTAAAGTCAAT 584
 |||||||
 Db 507 CCTCCTTCCAGCCCTCTGCTGGAGCAAGCTTCTAGTATGAGAGCTCTTCTTAAAGTCAAT 566
 585 GGAGAGGAAGCTTTGACAAATTAAGACAGAAACCGGATTTGCTGCAATAACCAATGCT 644
 |||||||
 Db 567 GGAGAGGAAGCTTTGACAAATTAAGACAGAAACCGGATTTGCTGCAATAACCAATGCT 626
 645 AATGATATGAGTGAAGAGAGCTACTAAAGAGATTTGGCAAGAAATTTGGTCTG 704
 |||||||
 Db 627 AATGATATGAGTGAAGAGAGCTACTAAAGAGATTTGGCAAGAAATTTGGTCTG 686
 705 TGCATTTCTGAATGTTCTTCTGTCMAACAGAGAAACATGAATCTGTCANATTTACAG 764
 |||||||
 Db 687 TGCATTTCTGAATGTTCTTCTGTCMAACAGAGAAACATGAATCTGTCANATTTACAG 746
 765 CTCTGATTTGCTCGAAGAGCAATGAGAGATTTGAGATTTATCAGAGTCTATGGTCTCA 824
 |||||||
 Db 747 CTCTGATTTGCTCGAAGAGCAATGAGAGATTTGAGATTTATCAGAGTCTATGGTCTCA 806
 825 AGTGAAGAGCACTTCTTTCACAGCAGATTCAGCCCAATCTGGAGAAATTTGCTGCGG 884
 |||||||
 Db 807 AGTGAAGAGCACTTCTTTCACAGCAGATTCAGCCCAATCTGGAGAAATTTGCTGCGG 866
 885 CATGGAATTAATCATCAGAAATCATCTTTTGGAGATTTCTGATTTGAG 947
 |||||||
 Db 867 CATGGAATTAATCATCAGAAATCATCTTTTGGAGATTTCTGATTTGAG 919
 RESULT 12
 AAS91688 standard; cDNA; 956 BP.
 ID AAS91688
 AC AAS91688;
 XX 13-FEB-2002 (first entry)
 DT
 DE DNA encoding novel human diagnostic protein #27492.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-0508631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX

DR WPI: 2001-639162/73.
 DR P-PSDB: AB627501.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1: SEQ ID NO 27492; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SO Sequence 956 BP; 265 A; 236 C; 243 G; 212 T; 0 other;
 Query Match 22.0%; Score 738.8; DB 23; Length 956;
 Best Local Similarity 93.7%; Pred. No. 6-7e-173;
 Matches 841; Conservative 0; Mismatches 12; Indels 15; Gaps 5;
 45 GCACAGAGCGGTAGACCCCTCTCTCTAAGTGGGAGGAGGAGCGGACGCAATTTTCA 104
 |||||||
 Db 67 GCACAGAGCGGTAGACCCCTCTCTCTAAGTGGGAGGAGGAGCGGACGCAATTTTCA 126
 105 CCTGTCGGGACAGCAACAGCACCATCTCTTGGGAGAACCTCTCCCTCTCTTGAAGAG 164
 |||||||
 Db 127 CCTGTCGGGACAGCAACAGCACCATCTCTTGGGAGAACCTCTCCCTCTCTTGAAGAG 186
 165 AAGATGTCGAATGGGTATTCACAGAGAGAAATTCGATATCATCTCTGTTTCAG 224
 |||||||
 Db 187 AAGATGTCGAATGGGTATTCACAGAGAGAAATTCGATATCATCTCTGTTTCAG 246
 225 GGCCAGGTTGAATATGATATCCAGAGTGGAGCCCTGTGCTGAGTACCTGACCTTTCTGCC 284
 |||||||
 Db 247 GGCCAGGTTGAATATGATATCCAGAGTGGAGCCCTGTGCTGAGTACCTGACCTTTCTGCC 268
 285 TGCAGAGTGAAGAGAGAGATTCAGAGAGAGTGGCCACCTCCGGGAACATCTAAGCACT 344
 |||||||
 Db 269 --CAGAGGTGAAGAGAGAGATTCAGAGAGAGTGGCCACCTCCGGGAACATCTAAGCACT 326
 345 TGAAGTCTGCTGCTGAGCAGCTTGGAGAGAGGAGTCTGGACCTTGGTGGTCTGGGAAAT 404
 |||||||
 Db 327 TGAAGTCTGCTGCTGAGCAGCTTGGAGAGAGGAGTCTGGACCTTGGTGGTCTGGGAAAT 386
 405 CGTGGAGGCGCTCCGGAGAACCGGACCGCTCTGGCCCGCCGCTACATCTTCCCTGAGCT 464
 |||||||
 Db 387 CGTGGAGGCGCTCCGGAGAACCGGACCGCTCTGGCCCGCCGCTACATCTTCCCTGAGCT 416
 465 CACGAGCTTCCCTCCAGCTGTTTGAAGAACGCTCATGATGATATCTCACTGCTGAA 523
 |||||||
 Db 447 CACGAGCTTCCCTCCAGCTGTTTGAAGAACGCTCATGATGATATCTCACTGCTGAA 506
 524 ACCTCTTCAGAGCC--ACTGTGGTGAAGCAAGCTTCTAGTGAAGAGCTCTTAAAGTGG 581
 |||||||
 Db 507 ACCTCTTCAGAGCCACCTCTGGGGGAGCAAGCTCTAGTTAAAGAGCTGTGATTAAGTG 566

| QY | 582 | CATGAGAGAGAACTGTTGGCAATTGGAAGACAGAAACCGGATGTGCTC | CGAAAAAGCA | 641 |
|-----------|---|---|-------------|-----|
| Db | 567 | CATGAGAGAGAACTGTTGGCAATTGGAAGACAGAAACCGGATGTGCTC <th>TGAAAAAGCA</th> <td>626</td> | TGAAAAAGCA | 626 |
| QY | 642 | TGAAAAATGAATCAGCTGTAAAGAGCTACTAAAAAGGATTTGTCCAGAAAT | AAAA-CTGGT | 700 |
| Db | 627 | TGAAAAATGAATCAGCTGTAAAGAGCTACTAAAAAGGATTTGTCCAGAAAT | AAAA-CTGGT | 686 |
| QY | 701 | TCTCTGATTTTCTGAATGTTCTTCTGTCGTCAAACAGGAAA-CATGAGACTG | CTAAAGAGTTA | 759 |
| Db | 687 | TCTCTGATTTTCTGAATGTTCTTCTGTCGTCAAACAGGAAA-CATGAGACTG | CTAAAGAGTTA | 746 |
| QY | 760 | ACAGGCTCTGATTTCTCAGAAAGCAATGAGAGATTGACAAATTTATCAATG | AGTTGATGGT | 819 |
| Db | 747 | ACAGGCTCTGATTTCTCAGAAAGCAATGAGAGATTGACAAATTTATCAATG | AGTTGATGGT | 806 |
| QY | 820 | CCTCAGAGTGAAGAGCAACTCTTTTCAACCCAGTTCAGCCAAATCTGTAAGAGAGTC | 879 | |
| Db | 807 | CCTCAGAGTGAAGAGCAACTCTTTTCAACCCAGTTCAGCCAAATCTGTAAGAGAGTC | 866 | |
| QY | 880 | TGGGGCATTGGAGATTAATCATCATGAAATATCTTTTGCAGATTTCTCTGATTTTCAG | 937 | |
| Db | 867 | TGGGGCATTGGAGATTAATCATCATGAAATATCTTTTGCAGATTTCTCTGATTTTCAG | 924 | |
| RESULT 13 | | | | |
| ID | AS01150 | standard; DNA: 3692 BP. | | |
| AC | AS01150: | | | |
| DT | 12-SEP-2001 | (first entry) | | |
| DE | Interferon induced nucleic acid, IFN5. | | | |
| XX | Interferon induced nucleic acid, autolimmune disease; lupus erythematosus | | | |
| KM | immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS; | | | |
| KM | graft rejection; viral infection; hepatitis; aplastic anemia; cancer; | | | |
| KM | human immunodeficiency virus; HIV; immune-mediated glomerulonephritis; | | | |
| KM | hematologic disease; chronic neutropenia; myocardial infarction; | | | |
| KM | neurological disease; Alzheimer's disease; Parkinson's disease; tumour; | | | |
| KM | amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5; ds. | | | |
| XX | Homo sapiens. | | | |
| OS | | | | |
| FT | Key | Location/Qualifiers | | |
| FT | CDS | 1..804 | | |
| FT | | /*tag- a | | |
| FT | | /product- *IFN5* | | |
| FT | | /note- *Interferon induced polypeptide* | | |
| XX | MO200118208-A2. | | | |
| XX | 15-MAR-2001. | | | |
| PD | | | | |
| XX | 08-SEP-2000; 2000MO-US24704. | | | |
| PF | | | | |
| XX | 08-SEP-1999; 99US-0152921. | | | |
| PR | 20-OCT-1999; 99US-0160575. | | | |
| PR | 20-JAN-2000; 2000US-0177104. | | | |
| PR | 07-SEP-2000; 2000US-0656633. | | | |
| XX | | | | |
| PA | (CURA-) CURAGEN CORP. | | | |
| PA | (BIOL) BIOGEN INC. | | | |
| XX | | | | |
| PI | Peyman JA, Da Silva A, Hochman P, Hsu A; | | | |
| XX | | | | |
| DR | WPI: 2001-235201/24. | | | |
| DR | P-PSDB: AAU00297. | | | |
| XX | | | | |
| PT | New interferon induced polypeptides and polynucleotide, useful for the diagnosis, prevention and treatment of immunological, cell | | | |

PT Proliferative disorders, such as lupus erythematosus, cancer, stroke
PT and Alzheimer's disease

PS Claim 9; Page 33-39; 13app; English.

XX The sequence represents interferon induced nucleic acid, IFN γ , IFN
XX nucleic acids and polypeptides are useful for treating or preventing a
XX pathology associated with IFN polypeptide in a human. They are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC acids, polypeptides and antibodies are useful for diagnosis, prevention
CC or treatment of variety of immunological and cell proliferative
CC disorders, such as autoimmune diseases e.g., lupus erythematosus,
CC immunodeficiency diseases such as acquired immunodeficiency syndrome
CC (AIDS), graft rejection, viral infections including hepatitis and human
CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC haematologic diseases such as aplastic anaemia and chronic neutropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon agonists, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans.

XX Sequence 3692 BP; 1166 A; 709 C; 806 G; 1011 T; 0 other;

SQ

Query Match 19.7%; Score 663; DB 22; Length 3092;
Best Local Similarity 92.9%; Pred. No. 8.2e-154;
Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1

45 GCACAGAGCGGTAGACCCTGCTTCTCTTAGTGGGGACGGGACAGCGGCAGCAGCAATTCGA 104
|||||
Db 55 GCACAGAGCGGTAGACCCTGCTTCTCTTAGTGGGGACGGGACAGCGGCAGCAGCAATTCGA 114

105 CCTGTCCGCCAGACAACAGCACCATCTGCTTTGGAGAACCTCTCCCTCTCTGAGAAG 164
|||||
Db 115 CCTGTCCGCCAGACAACAGCACCATCTGCTTTGGAGAACCTCTCCCTCTCTGAGAAG 174

165 AAAGATGTGAAATGGATTATTCACAGAGAGAAATTCGGCTATTCATCTCTTTCAG 224
|||||
Db 175 AAAGATGTGAAATGGATTATTCACAGAGAGAAATTCGGCTATTCATCTCTGAG 234

225 GGCCAGGAGTGAATAATTAATCACTCAGGTGAGACCTGTGTGACTACTGACTCTTGCC 284
|||||
Db 235 GGCCAGGAGTGAATAATTAATCACTCAGGTGAGACCTGTGTGACTACTGACTCTTGCC 296

285 TGACAGAGCTGAAGAGACAGATTTCAGAGACAGTCCGACCTCCGGAGACATATTAAGT 344
|||||
Db 257 --CAGAGGTGAAGAGACAGATTTCAGAGACAGTCCGACCTCCGGAGACATATTAAGT 314

345 TGAAGTCTGCTGAGCACCCTTGGAGAGAGAGTCTGGACCTTGGTTGGACTGCGAATT 404
|||||
Db 315 TGAAGTCTGCTGAGCACCCTTGGAGAGAGAGTCTGGACCTTGGTTGGACTGCGAATT 374

405 CGTGGAGGCGCTCCGGAGAACCGGCGACCCCTCGGCGCGCGCGGTACATGATCTGACG 464
|||||
Db 375 CGTGGAGGCGCTCCGGAGAACCGGCGACCCCTCGGCGCGCGCGGTACATGATCTGACG 434

465 CACGAGCTGACCTTCATCTGATTTAGAGACGCTCATGATGATATCTCCAATTTTAA 524
|||||
Db 435 CACGAGCTGACCTTCATCTGATTTAGAGACGCTCATGATGATATCTCCAATTTTAA 494

525 CCTCCTTACGCCCACTCGTGTGGACAGCTTTAGTTAAGACGCTTGGAAAATTCAT 584
|||||
Db 495 CCTCCTTACGCCCACTCGTGTGGACAGCTTTAGTTAAGACGCTTGGAAAATTCAT 554

585 GGAGAGGAGAACTGTGACAATTCAGACAGAAAAACCGGATTGCTGCTGCAGAAAAAATGG 644
|||||
Db 555 GGAGAGGAGAACTGTGACAATTCAGACAGAAAAACCGGATTGCTGCTGCAGAAAAAATGG 614

| Query Match | Best Local Similarity | Score | DB | Length | Matches | Conservative | Indels | Gaps |
|--|--|---------|--------|--------|---------|--------------|--------|------|
| 19.7% | 92.9% | 663 | 22 | 4701 | 722 | 0 | 15 | 1 |
| <p>disorders/such as autoimmune diseases e.g. lupus erythematosus, immunodeficiency diseases such as acquired immunodeficiency syndrome (AIDS), graft rejection, viral infections including hepatitis and human immunodeficiency virus (HIV), immune-mediated glomerulonephritis, haematologic diseases such as aplastic anaemia and chronic leucopenia and cancer. In addition they are also useful for treating or diagnosing various disorders associated with cell death, including myocardial infarction, stroke, neurological diseases including Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular atrophy. IFN nucleic acids and polypeptides are also useful for identifying interferon-like proteins and interferon agonists, for screening drugs and compounds which inhibit or enhance IFN activity or function and as targets for the identification of small molecules that are immunostimulatory, immunosuppressive, or stimulate or suppress natural cell or tumour cell growth in mammals, including humans.</p> | | | | | | | | |
| Sequence | 3704 BP; | 1168 A; | 713 C; | 809 G; | 1014 T; | 0 other; | | |
| 45 | GCACAGAGCGGTAGACCCCTGCTTCTCTAGTGGGACGGGACAGCGGACGATATTTCA | 104 | | | | | | |
| 67 | GACAGAGCGGTAGACCCCTGCTTCTCTAGTGGGACGGGACAGCGGACGATATTTCA | 126 | | | | | | |
| 105 | CCTGTCGCGACACACAGACACACCTCTGCTTGGAGAACCTCTCCCTCTCTATATAAG | 164 | | | | | | |
| 127 | CCTGTCGCGACACACACACACCTCTGCTTGGAGAACCTCTCCCTCTCTATATAAG | 186 | | | | | | |
| 165 | AAAGATGTGATGGGTATTCACAGACAGAGATTTCCCTATCTCATCTCTGTTTCAAG | 224 | | | | | | |
| 187 | AAAGATGTGATGGGTATTCACAGACAGAGATTTCCCTATCTCATCTCTGTTTCAAG | 246 | | | | | | |
| 225 | GGCCAGGCGTAAATGTATACATCCAGGTGAGGCGCTGTGTGACTACCTCACTCTGCGC | 284 | | | | | | |
| 247 | GGCCAGGCGTAAATGTATAC----- | 268 | | | | | | |
| 285 | TGCAGAGGTGAAGAGACAGATTCAGAGACAGTCCGACACTCCGGAGACATATAAGT | 344 | | | | | | |
| 269 | --CAAGAGGTGAAGAGACAGATTCAGAGACAGTCCGACACTCCGGAGACATATAAGT | 326 | | | | | | |
| 345 | TGAACGTGCTGCTGAGACACTTTGGAGAAAGAGACTGTGGACCTTGGTTGGACAGTAAAT | 404 | | | | | | |
| 327 | TGAACGTGCTGCTGAGACACTTTGGAGAAAGAGACTGTGGACCTTGGTTGGACAGTAAAT | 386 | | | | | | |
| 405 | CGTGAAGCGCTCCGGAGAACCGGGACGCCCTGTGGCGCGCCGCTACATGAATCTTACGT | 464 | | | | | | |
| 387 | CGTGAAGCGCTCCGGAGAACCGGGACGCCCTGTGGCGCGCCGCTACATGAATCTTACGT | 446 | | | | | | |
| 465 | CACGAGACTGGCCCTCCATTCGTTTGAAGACCTCATATGATATCTCCAAATCTTGAA | 524 | | | | | | |
| 447 | CACGAGACTGGCCCTCCATTCGTTTGAAGACCTCATATGATATCTCCAAATCTTGAA | 506 | | | | | | |
| 525 | CCTCCTTACGCCACTCTGTGGAGAAAGCTTGTAGTAAAGACGCTTGGATTAATGAT | 584 | | | | | | |
| 507 | CCTCCTTACGCCACTCTGTGGAGAAAGCTTGTAGTAAAGACGCTTGGATTAATGAT | 566 | | | | | | |
| 585 | GGAGAGGAGACTGTGTACATTTGAACACAGAACCGGATTTGCTGTGCGAATAATG | 644 | | | | | | |
| 567 | GGAGAGGAGACTGTGTACATTTGAACACAGAACCGGATTTGCTGTGCGAATAATG | 626 | | | | | | |
| 645 | AAATGAATGAGTGAAGAGACTCTTAAAGAGATTGTGCGAATAAGAAATGCTTTC | 704 | | | | | | |
| 627 | AAATGAATGAGTGAAGAGACTCTTAAAGAGATTGTGCGAATAAGAAATGCTTTC | 686 | | | | | | |
| 705 | TGCATTTTGCATGTCTTCTGTCAAACAGAAACATGACTGTGTCAGATTTTAAACAG | 764 | | | | | | |
| 687 | TGCATTTTGCATGTCTTCTGTCAAACAGAAACATGACTGTGTCAGATTTTAAACAG | 746 | | | | | | |
| 765 | CTCTGATTTGCTAGAAAGCAATGACAGACTTGAAGATTTATACAAAGTTGATGCTG | 821 | | | | | | |
| 747 | CTCTGATTTGCTAGAAAGCAATGACAGACTTGAAGATTTATACAAAGTTGATGCTG | 803 | | | | | | |

RESULT 15

AAH3745 standard: CDNA: 609 BP.

AAH3745:

03-SEP-2001 (first entry)

Human colon cancer antigen encoding CDNA SEQ ID NO: 801.

Human colon cancer; colon cancer antigen: diagnosis: detection;

colorectal carcinoma; ss.

Homo sapiens.

WO200122920-A2.

05-APR-2001.

28-SEP-2000; 2000MO-US26524.

29-SEP-1999; 99US-0157137.

03-NOV-1999; 99US-0163280.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA;

WPI: 2001-235357/24.

P-PSDB: AAG74314.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers.

Claim 1: Page 2766; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 609 BP; 229 A; 90 C; 104 G; 174 T; 12 other;

Query Match 17.2% Score 577.2; DB 22; Length 609;

Best Local Similarity 98.6%; Pred No. 6.5e-133;

Matches 573; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

2785 ATTTGGAATTACAGATGCAAGTATATGAAAGAAATGAAACCAAGAAATAT 2844

5 ATTTGGAATTACAGATGCAAGTATATGAAAGAAATGAAACCAAGAAATAT 64

2845 GCCAAGCATTTACAGAAATACCATCATTAATACCTTCTTGGCAAAATGAGAGTGTG 2904

65 GCCAGCATTTACAGAAATACCATCATTAATACCTTCTTGGCAAAATGAGAGTGTG 124

2905 CTAGCTGTCTGGGAGATATCATTAATTGAGAAATGATGATGATTAATATGACC 2964

DB 125 CTAGCTGTCTGGGAGATATCATTAATTGAGAAATGATGATGATTAATATGACC 184

QY 2965 CCAGATTTCAAGGAATCTTACATTTGTAGAGAAAACAGACTGCAAAACAGAGTGC 3024

DB 185 CCAGATTTCAAGGAATCTTACATTTGTAGAGAAAACAGACTGCAAAACAGAGTGC 244

QY 3025 GACTATCAAAATTAATGTTGAAATCATCTGCAAAATGTCGAGGCTGGGGAAATGATG 3084

DB 245 GACTATCAAAATTAATGTTGAAATCATCTGCAAAATGTCGAGGCTGGGGAAATGATG 304

QY 3085 GTGCACAAAGGCTTAGATTTGCTGCTCAAAATAGGAATTTGTAGTCTTCAAA 3144

DB 305 GTGCACAAAGGCTTAGATTTGCTGCTCAAAATAGGAATTTGTAGTCTTCAAA 364

QY 3145 AATAATTCACAAAGAAACAAATACAAAAAGTGGTAGAATTACTATCAATTCAT 3204

DB 365 AATAATTCACAAAGAAACAAATACAAAAAGTGGTAGAATTACTATCAATTCAT 424

QY 3205 CTGACTATTCAGATGCTGTTTATTAGTATGATGAGATTAAGCACTGATTCAT 3264

DB 425 CTGACTATTCAGATGCTGTTTATTAGTATGATGAGATTAAGCACTGATTCAT 484

QY 3265 TTTAAATACATCATGTTAAACATTTAATGATTAATGATTAATGATTAATGATTA 3324

DB 485 TTTAAATACATCATGTTAAACATTTAATGATTAATGATTAATGATTAATGATTA 544

QY 3325 AGAAGTACATTAAGATCAATTAATTAATGATTTACTCTG 3365

DB 545 AGAAGTACATTAAGATCAATTAATTAATGATTTACTCTG 585

Search completed: May 16, 2003, 19:52:12
 Job time: 744 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using SW model

Run on: May 16, 2003, 12:54:36 ; Search time 8627 seconds

11351.680 Million cell updates/sec

Title: US-09-515-363C-1

Sequence: 1 gcgcgcgcgcctgagagccc.....aaatgattgt | ractctg 3365

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 410,330

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Maximum DB seq length: 200000000000

| Pre-processing: | Minimum Match | Maximum Match |
|------------------|------------------|--------------------|
| Post-processing: | Minimum Match 0% | Maximum Match 100% |

| | |
|----------------------------|------|
| Maximum Match | 100% |
| Listing first 45 summaries | |

Database

| | |
|-----|----------------|
| 1: | gb.ba.* |
| 2: | gb.htg.* |
| 3: | gb.in.* |
| 4: | gb.om.* |
| 5: | gb.ov.* |
| 6: | gb.pat.* |
| 7: | gb.ph.* |
| 8: | gb.pl.* |
| 9: | gb.pr.* |
| 10: | gb.ro.* |
| 11: | gb.sts.* |
| 12: | gb.sy.* |
| 13: | gb.un.* |
| 14: | gb.vl.* |
| 15: | em.ba.* |
| 16: | em.fun.* |
| 17: | em.hum.* |
| 18: | em.in.* |
| 19: | em.mu.* |
| 20: | em.om.* |
| 21: | em.ov.* |
| 22: | em.ov.* |
| 23: | em.pat.* |
| 24: | em.ph.* |
| 25: | em.pl.* |
| 26: | em.ro.* |
| 27: | em.sts.* |
| 28: | em.un.* |
| 29: | em.vl.* |
| 30: | em.hg.hum.* |
| 31: | em.hg.in.v.* |
| 32: | em.hg.other.* |
| 33: | em.hg.mus.* |
| 34: | em.hg.pln.* |
| 35: | em.hg.rod.* |
| 36: | em.hg.nam.* |
| 37: | em.hg.vlt.* |
| 38: | em.sic.* |
| 39: | em.hgo.hum.* |
| 40: | em.hgo.mus.* |
| 41: | em.hgo.other.* |

Pred. No. is the number of results predicted by chance to have a

| Result No. | Score | Query Match | length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|---------------------|
| 1 | 3365 | 100.0 | 3380 | 9 | AF095844 | AF095844 Homo sapi |
| 2 | 3338.8 | 99.2 | 3372 | 6 | AA300832 | AA300832 Sequence |
| 3 | 3338.8 | 98.2 | 3373 | 9 | AY017378 | AY017378 Homo sapi |
| 4 | 2202.4 | 65.5 | 3771 | 10 | AF374384 | AF374384 Homo sapi |
| 5 | 2135.4 | 63.5 | 3078 | 10 | AY075132 | AY075132 Homo sapi |
| 6 | 1766 | 52.5 | 1776 | 9 | AK056293 | AK056293 Homo sapi |
| 7 | 1435.6 | 42.7 | 1443 | 6 | AX300841 | AX300841 Sequence |
| 8 | 1402 | 41.7 | 2468 | 10 | BC025508 | BC025508 Mus muscu |
| 9 | 1254.6 | 37.3 | 1284 | 6 | AA300838 | AA300838 Sequence |
| 10 | 1194.4 | 35.5 | 1213 | 6 | BC007966 | BC007966 Homo sapi |
| 11 | 1166 | 35.2 | 2050 | 10 | BC004031 | BC004031 Mus muscu |
| 12 | 803 | 23.9 | 1258 | 6 | AX098232 | AX098232 Sequence |
| 13 | 803 | 23.9 | 1270 | 6 | AX098236 | AX098236 Sequence |
| 14 | 663 | 19.7 | 3692 | 6 | AA098234 | AA098234 Sequence |
| 15 | 663 | 19.7 | 3704 | 6 | AA098238 | AA098238 Sequence |
| 16 | 623.2 | 18.5 | 95417 | 9 | AC010876 | AC010876 Homo sapi |
| 17 | 384 | 11.4 | 923 | 10 | BC019605 | BC019605 Mus muscu |
| 18 | 371.2 | 11.0 | 392 | 6 | AX071769 | AX071769 Sequence |
| 19 | 356.8 | 10.6 | 387 | 6 | AX071783 | AX071783 Sequence |
| 20 | 303 | 9.0 | 16381 | 9 | AC007750 | AC007750 Homo sapi |
| 21 | 285.4 | 8.5 | 301 | 6 | 150897 | 150897 Sequence |
| 22 | 284.4 | 8.5 | 118891 | 2 | AC094719 | AC094719 Rattus no |
| 23 | 284.4 | 8.5 | 182906 | 2 | AC111919 | AC111919 Rattus no |
| 24 | 279.4 | 8.3 | 210317 | 2 | AC115074 | AC115074 Mus muscu |
| 25 | 269.2 | 8.0 | 377 | 6 | A74554 | A74554 Sequence |
| 26 | 269.2 | 8.0 | 377 | 6 | A77533 | A77533 Sequence |
| 27 | 223.8 | 6.6 | 2326 | 10 | AF316999 | AF316999 Mus muscu |
| 28 | 221.4 | 6.6 | 2336 | 10 | BC022209 | BC022209 Mus muscu |
| 29 | 179.6 | 5.3 | 2584 | 9 | BC014949 | BC014949 Homo sapi |
| 30 | 178 | 5.3 | 2613 | 9 | AC021416 | AC021416 Homo sapi |
| 31 | 172 | 5.1 | 182906 | 2 | AC111919 | AC111919 Rattus no |
| 32 | 166.8 | 5.0 | 118891 | 2 | AC094719 | AC094719 Rattus no |
| 33 | 153 | 4.5 | 164 | 6 | AA300834 | AA300834 Sequence |
| 34 | 150.2 | 4.5 | 2250 | 9 | AK097669 | AK097669 Homo sapi |
| 35 | 148.6 | 4.4 | 3265 | 9 | AF038963 | AF038963 Homo sapi |
| 36 | 141.8 | 4.2 | 160 | 6 | A74468 | A74468 Sequence |
| 37 | 141.8 | 4.2 | 160 | 6 | A77447 | A77447 Sequence |
| 38 | 130 | 3.9 | 112361 | 9 | AC027316 | AC027316 Homo sapi |
| 39 | 130 | 3.9 | 143200 | 9 | AC008413 | AC008413 Homo sapi |
| 40 | 124.6 | 3.7 | 193 | 6 | 150896 | 150896 Sequence |
| 41 | 114 | 3.4 | 4024 | 4 | AF181119 | AF181119 Sus scro |
| 42 | 109.6 | 3.3 | 2633 | 9 | HSMB02366 | HSMB02366 Homo sapi |
| 43 | 109.6 | 3.3 | 3026 | 9 | AC023661 | AC023661 Homo sapi |
| 44 | 72.4 | 2.2 | 100635 | 9 | AC003104 | AC003104 Homo sapi |
| 45 | 72.4 | 2.2 | 154808 | 2 | AC105024 | AC105024 Homo sapi |

ALIGNMENTS

| | | | | | |
|------------|---|---------|------|--------|-----------------|
| RESULT 1 | | | | | |
| LOCUS | AF095844 | 3380 bp | mRNA | linear | 161 23-JAN-2002 |
| DEFINITION | Homo sapiens melanoma differentiation associated protein-5 (MDA5) | | | | |
| ACCESSION | AF095844 | | | | |
| VERSION | AF095844.1 | | | | |
| KEYWORDS | AF095844.1 GI:1344593 | | | | |
| SOURCE | Homo sapiens. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | | | |
| AUTHORS | Kang,D.C., Gopalakrishnan,R.V., Wu,Q., Jankowsky,E., Pye,A.M. and Fisher,P.B. | | | | |

TITLE mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 637-642 (2002)
MEDLINE 21664412
PUBMED 11805321
REFERENCE 2 (bases 1 to 3380)
AUTHORS Kang, D.-C. and Fisher, P. B.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1998) Urology, Columbia University, P&S, 630 West 168th, New York, NY 10032, USA

FEATURES
source location/Qualifiers
1..3380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HO-1"
/tissue_type="melanoma"
1..3380
/gene="MDA5"
/gene="MDA5"
169..3246
/codon_start=1
/product="melanoma differentiation associated protein-5"
/protein_id="AF34368.1"
/db_xref="GI:11344594.1"
/translation="MSGSTDENRFLVLSCEPRARVKKHIGVEPYLDYITFLPEVKE
QIQRTVATSGNMAVELLSTLEKGVHMGTRVEVAKIGTSHLARYLNPELTDL
PSPFENAHDEYDLDLNLDTLVKDLKLVYDLKCMFELTITIDRRKIAAENGN
ESGVELLRKIVQKEMNPSAFNLVLRGTNNELVELTSPCESNAELENLSVDGP
GVEEQLSTTVQPNLEKVEVMGEMNNSPADSDPCCDEDEDIKKPKLIDEDRLMLFPE
NSNMGSDGTGSDSDENVAARASPEEQLDRPOMVAPALPGRKLLITCLPTGSC
KTRAVVYIAKDLHDKKASBPQKIVLVNLYVLEQLKPEQFFLKKMTRVIGLGS
DYLKISPEYVAKSCDIIITQQLIENSLNLENGEDAVLDSFLIIDCHHTNK
EAVYNNIRHRYLMOKLNKRLKENKPYIPQLIGLTSVGVGATQQAEEHLK
LCANLDAFTIKTKENLDOLKNOIOEPCKEALADATRPFEKLELIMRIQYCG
MSPMDFEPTQPEQWAIOMEKKAQKGRKRYCAEHLKYNALQINDITIMIDAT
HLETFYNEKKEFAVIEDSDSGDDPCDDEDEDIKKPKLIDEDRLMLFPE
NNKMLKRLAENREYENELTKIRNTIMQYTRTESARLIPTKIRGSAYVLSQWITE
NEKRAEVCYKAHLIGAGHSEFPKPTONEKREVSRTKIKINLLIATVAEEGLDI
KECNTVIRYGLTNEIAVQAGARARAEKSTYVLAHSNVIHEIETVNDREKMKYK
AHCYQNMKPREYAHKILEQNASIMERKMTKRMIAKYNNSLITFLKNCISVLA
CSGEDIHYLEKKNHYNMTPEFKELYIVENKALDOKCAIYQINGEILICCGCAQTMH
VHKGLDCEIKRNFVYVFKNNSTKQYKNWELPITFPMIDYSCECLFSDSD"

BASE COUNT 1153 a 644 c 753 g 830 t
ORIGIN

Query Match 100.0%; Score 3365; DB 9; Length 3380;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 361 ACCCTGGAGAAAGAGTGTGACACCTTGTTGACCTGGGAATTCGTGGAGAGTTCGGG 420
DB 361 ACCCTGGAGAAAGAGTGTGACACCTTGTTGACCTGGGAATTCGTGGAGAGTTCGGG 420
OY 421 AGAACCGGAGCCCTCTGTGGCCCGCCGTACATGAACCTGTAGCTCAGGACTTGTCT 480
DB 421 AGAACCGGAGCCCTCTGTGGCCCGCCGTACATGAACCTGTAGCTCAGGACTTGTCT 480
OY 481 CCATGTTTGAAGAACGCTCATGATGATATCTCCAACTGTAAGCTCTCTCAGTCCACT 540
DB 481 CCATGTTTGAAGAACGCTCATGATGATATCTCCAACTGTAAGCTCTCTCAGTCCACT 540
OY 541 CTGTGTGACAAAGCTTCTAGTAGAGAGCTTGGATTAAGTGCATGAGAGAGAAATGTTG 600
DB 541 CTGTGTGACAAAGCTTCTAGTAGAGAGCTTGGATTAAGTGCATGAGAGAGAAATGTTG 600
OY 601 ACAATTGAAGACAGAAACCGGATTCTCTGTGCAAAAACATGGAAATGAATCAATGTGA 660
DB 601 ACAATTGAAGACAGAAACCGGATTCTCTGTGCAAAAACATGGAAATGAATCAATGTGA 660
OY 661 AGAGAGCTACTAAAGAGATTGTGAGAAAGAAACCTGCTCTGCATTTCTGTGATGTT 720
DB 661 AGAGAGCTACTAAAGAGATTGTGAGAAAGAAACCTGCTCTGCATTTCTGTGATGTT 720
OY 721 CTTCGTCAAACAGAGAAACATGAACCTTGTCCAGAGTTAACAGGCTGTGATGCTTACGA 780
DB 721 CTTCGTCAAACAGAGAAACATGAACCTTGTCCAGAGTTAACAGGCTGTGATGCTTACGA 780
OY 781 AGCAATGAGAGATTGAAATTAATACAACTTGTGCTCTCAAGTGAAGAGAACTT 840
DB 781 AGCAATGAGAGATTGAAATTAATACAACTTGTGCTCTCAAGTGAAGAGAACTT 840
OY 841 CTTTCAACACAGTTCAGCAAAATGTGAGAAAGAGTCTGGGCGATGAGAAATTAATCA 900
DB 841 CTTTCAACACAGTTCAGCAAAATGTGAGAAAGAGTCTGGGCGATGAGAAATTAATCA 900
OY 901 TCAGAAATATCTTTTGCAGATTCTTCTAGTTTCAAGATCAGACACAGTTTGACAGAA 960
DB 901 TCAGAAATATCTTTTGCAGATTCTTCTAGTTTCAAGATCAGACACAGTTTGACAGAA 960
OY 961 GGAATGTCAGCTGTTTGAATGAAGAGCTTGTGACATTAACACCAATGAGGAGTATTCA 1020
DB 961 GGAATGTCAGCTGTTTGAATGAAGAGCTTGTGACATTAACACCAATGAGGAGTATTCA 1020
OY 1021 GGCACCATAGGAGAGTATTCAGATGAGAGATGTGGCAGCAGAGCATCCCGTAGCA 1080
DB 1021 GGCACCATAGGAGAGTATTCAGATGAGAGATGTGGCAGCAGAGCATCCCGTAGCA 1080
OY 1081 GAATTCACAGCTCAGGCTTACCAAAATGAAGTTGCCAGCCAGGCTTGGAAAGGAGCAAT 1140
DB 1081 GAATTCACAGCTCAGGCTTACCAAAATGAAGTTGCCAGCCAGGCTTGGAAAGGAGCAAT 1140
OY 1141 ATCATCATCTGCTCCCTACAGGGAGGTGGAAGAAACCAAGTGGCTTATCAATTCAGAG 1200
DB 1141 ATCATCATCTGCTCCCTACAGGGAGGTGGAAGAAACCAAGTGGCTTATCAATTCAGAG 1200
OY 1201 GATCACTTAGACAAAGAAAGAAACATCTGAGCTGTGAAAGTATAGTTTCTTCAAT 1260
DB 1201 GATCACTTAGACAAAGAAAGAAACATCTGAGCTGTGAAAGTATAGTTTCTTCAAT 1260
OY 1261 AAGGTACTGTAGTGAACAGCTCTCCGCAAGAGAGTTCCAACTTTTGAAGAAATGG 1320
DB 1261 AAGGTACTGTAGTGAACAGCTCTCCGCAAGAGAGTTCCAACTTTTGAAGAAATGG 1320
OY 1321 TATGCTGTTATTTGAATTAAGTGGTATACCAACTGAAATATATATTTCCAAATTTGTC 1380
DB 1321 TATGCTGTTATTTGAATTAAGTGGTATACCAACTGAAATATATATTTCCAAATTTGTC 1380
OY 1381 AAGTCTGTGATATATATAGTACAGCTCAAAATCTTGAAGTCCCTTAAAGCTTG 1440
DB 1381 AAGTCTGTGATATATATATAGTACAGCTCAAAATCTTGAAGTCCCTTAAAGCTTG 1440

TITLE Rh16 polypeptides and its fragments and polynucleotides encoding
said polypeptides and therapeutic uses
JOURNAL Patent: WO 0185955-A1 15-NOV-2001;
Istac (FR) : INSTITUT PASTEUR DE LILIE (FR)

FEATURES

Location/Qualifiers

1..3372

/organism="Homo sapiens"

CDS

155..3332

/note="unnamed protein product"

/protein_id="CAD13049.1"

/db_xref="GI:1738211"

/translation="MSGSTDENFRVLIISCFRAVYKMIIVEPVLYLITFLPAEYKE
QIQTAVATSGNMAVELLSTLEKVMHGMTRVEVAPRTGSLARVYNPELIDL
PSPFENAHDEYQLNLOPLVDKLVLRVLDKCMERLITLIEDNRILAAEENG
ESGVELLRIYOKENMFSAFLNLRGTNGELVDELTPNCSSEANLEINLSVDGP
OYEBOLSTTVQPNLEKYMENSSSPADSVSVSHITSLAEQSVSLDESLG
NSNMSSDSGTMSDSDENVAARPELDRIPYQMEVAPALRGKNIITCLPTGSG
KTRAVVITAKDLKKRASEPQVIVLVNVLVEQLVPEEPFELKMYRVLGLSG
DTQLKISPEPVVSCDIIISTQILENSLLENEDLVOLDSFPIIDECHEITNK
EAVYNNIMHLYLMOKLNNRLKKNKPVIPYQILGLTASGVGATQQAEEHILK
LCANDAFITIKVENLDOLKNOIOEPCKKFAIDATPRIPKEXLEIPMTRIOTYCO
MSPMDPTGPPEOMAIOMEKAAEGRRKSVCAHHLKINKALQINDTIRMIDAT
HLETPYNEKOKKRAVIDEDSDSDSDYCCDDEDEDKPKPLKIDETRLMTLFE
NKKILKRLAENPEIENKILKRLNTIMPOVTRTESARLITKTRQSAVLSQWITE
NEKFAVEVKAHHLIGASHSEFPMTQNEQEVLSKPIINILITAVTAEBLDI
KECNIVIRGLATNEIAMOARARADESTVIVASHSVIEREIVNDRKEMMYK
AHCIVONKPREYAKHILELOMSIMEKMKTRKIAIKYNNPILITCKNSVLA
CSGEDIHYLEKHHYNTPEFEYKIVVENVALOKCAVNLNGELIICKQCAQMTM
VHKGLDLCIKRNRVYVFKNNSTKQYKKNVELPITELINDYSECLSPED"

BASE COUNT 1157 a 637 c 748 g 830 t
ORIGIN

Query Match 99.2%; Score 3338.8; DB 6; Length 3372;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3343; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

15 GAGCCCTGTGAGAACCTGCTATTTGTCAGGACAGAGCGGTAGACCTTCTCTTAG 74
1 GGGCCCTGTGAGAACCTGCTATTTGTCAGGACAGAGCGGTAGACCTTCTCTTAG 60
75 TGGGCGAGGAGAGCGGACAGACATTTACCTGTCCCGACAGACAGATTCATCTGCT 134
61 TGGGCGAGGAGAGCGGACAGACATTTACCTGTCCCGACAGACAGATTCATCTGCT 120
135 TGGGAGAACCTCTCCTCTCTGAGAGAGAGATGTGAATGGTATCTTACAGACA 194
121 TGGGAGAACCTCTCCTCTCTGAGAGAGAGATGTGAATGGTATCTTACAGACA 180
195 GAATTCGGCTATCTCATCTGCTGCTTCAGGGCGAGGATGAATAATGTA 254
181 GAATTCGGCTATCTCATCTGCTGCTTCAGGGCGAGGATGAATAATGTA 240
255 GGCTGTGTGAGACTACCTACCTTTTGGCTGAGAGAGTGAAGAGACAG 314
241 GGCTGTGTGAGACTACCTACCTTTTGGCTGAGAGAGTGAAGAGACAG 300
315 AGTGGCACCTCCGAGAGACATGAGGACAGTGAAGTGCCTGAGAGACCT 374
301 AGTGGCACCTCCGAGAGACATGAGGACAGTGAAGTGCCTGAGAGACCT 360
375 AGTGGCACCTGCTGTTGAGACTGCGGAATTCGTGAGAGAGTGCAGAGAT 434
361 AGTGGCACCTGCTGTTGAGACTGCGGAATTCGTGAGAGAGTGCAGAGAT 420
435 TGTGGCGCGCGCTACATGAACCTGTAGCTGACGAGACTTGGCTCTCCAT 494
421 TGTGGCGCGCGCTACATGAACCTGTAGCTGACGAGACTTGGCTCTCCAT 480
495 CGCTCATGATATCTTCAACTGCTGAACCTCTTCCGACCACTGTA 554
481 CGCTCATGATATCTTCAACTGCTGAACCTCTTCCGACCACTGTA 540

555 TCTAGTAGAGAGCTGTTGATTAAGTGCATGAGAGAGAGACTGTTGACATTA 614
541 TCTAGTAGAGAGCTGTTGATTAAGTGCATGAGAGAGAGACTGTTGACATTA 600
615 AAACCGGATTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 674
601 AAACCGGATTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
675 AAGGATTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 734
661 AAGGATTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
735 AAACGATGACCTTTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 794
721 AAACGATGACCTTTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
795 TGAGAAATTTATCACAAGTTGATGCTGCTGAGAGAGAGAGAGAGAGAG 854
781 TGAGAAATTTATCACAAGTTGATGCTGCTGAGAGAGAGAGAGAGAGAG 840
855 TCAGGCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 914
841 TCAGGCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
915 TCAGGCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974
901 TCAGGCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
975 CTTAGATGAAGTCTTGGACATTAACAGAGAGAGAGAGAGAGAGAGAGAG 1034
961 CTTAGATGAAGTCTTGGACATTAACAGAGAGAGAGAGAGAGAGAGAGAG 1020
1035 TGATTCAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1094
1021 TGATTCAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1095 GCCTTACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1154
1081 GCCTTACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
1155 CCCTACAG 1214
1141 CCCTACAG 1200
1215 GAG 1274
1201 GAG 1260
1275 TGACAG 1334
1261 TGACAG 1320
1335 ATTAAGTGTGATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
1321 ATTAAGTGTGATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
1395 TATTATCAGTACAGTCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
1381 TATTATCAGTACAGTCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
1455 TGTGGGTGTTCAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
1441 TGTGGGTGTTCAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1515 CAAAG 1574
1501 CAAAG 1560
1575 TAGACTCAAG 1634
1561 TAGACTCAAG 1620

| Db | 2701 | CGGAGAGGAAGATGATGCTATTAAGCATATCATTTGCTTCAAAATATGAAACCCAGATGAACTA | 2760 |
|------------|--|--|--------|
| QY | 2775 | TGCTCATAGATTTTGGAAATTACAGATGCAGAAAGTAAATGGAAGAAATGAAACCCAA | 2834 |
| Db | 2761 | TGCTCATAGATTTTGGAAATTACAGATGCAGAAAGTAAATGGAAGAAATGAAACCCAA | 2820 |
| QY | 2835 | GAGAAATTTGGCAAGCATTTACAGAAATTAACCCATCATTAATTAACCTTCTTGAAATA | 2894 |
| Db | 2821 | GAGAAATTTGGCAAGCATTTACAGAAATTAACCCATCATTAATTAACCTTCTTGAAATA | 2880 |
| QY | 2895 | CTGCAGTGTGCTAGCCTGTCTTGGGAAGATATCCATGTAATATGGAATATGCAACAGCT | 2954 |
| Db | 2881 | CTGCAGTGTGCTAGCCTGTCTTGGGAAGATATCCATGTAATATGGAATATGCAACAGCT | 2940 |
| QY | 2955 | CAATATGACCCCGAATTCAGGAACCTTACATTTGAAGAGAAACCAAGCACTTCAGAAA | 3014 |
| Db | 2941 | CAATATGACCCCGAATTCAGGAACCTTACATTTGAAGAGAAACCAAGCACTTCAGAAA | 3000 |
| QY | 3015 | GAAAGTGTCCGCACTTCAGAAATTAAGTGAATATCATCTGCAAAATGTGCCAGATTTGAG | 3074 |
| Db | 3001 | GAAAGTGTCCGCACTTCAGAAATTAAGTGAATATCATCTGCAAAATGTGCCAGATTTGAG | 3060 |
| QY | 3075 | AACATATGATGGTGCAACAAAGCTTGATTTGGCTGTGCTCAAAATTAAGAAATTTGAT | 3134 |
| Db | 3061 | AACATATGATGGTGCAACAAAGCTTGATTTGGCTGTGCTCAAAATTAAGAAATTTGAT | 3120 |
| QY | 3135 | GGTTTTCAGAAAATTAATTCAGCAAGAAACAAATATCAAAAAGTGGGTAGATTAACCTATC | 3194 |
| Db | 3121 | GGTTTTCAGAAAATTAATTCAGCAAGAAACAAATATCAAAAAGTGGGTAGATTAACCTATC | 3180 |
| QY | 3195 | ATTTCGCCAATCTTGACATATTCAGAAATGCTGTTATTATGATGATGAGATAGACCTTAT | 3254 |
| Db | 3181 | ATTTCGCCAATCTTGACATATTCAGAAATGCTGTTATTATGATGATGAGATAGACCTTAT | 3240 |
| QY | 3255 | TGAAGATCTTTTAAATATCTATCATGTTAAACATTAATATGATTAATGATTAATGATTA | 3314 |
| Db | 3241 | TGAAGATCTTTTAAATATCTATCATGTTAAACATTAATATGATTAATGATTAATGATTA | 3300 |
| QY | 3315 | ATTATGCTACAGACTGACATAGAAATCAATCAATTAATGATTTGTTTACTCT | 3364 |
| Db | 3301 | ATTATGCTACAGACTGACATAGAAATCAATCAATTAATGATTTGTTTACTCT | 3350 |
| RESULT 3 | | | |
| AY017378 | | 3373 bp | linear |
| LOCUS | | | |
| DEFINITION | Homo sapiens RNA helicase-DEAD box protein RH16 mRNA, complete cds. | | |
| ACCESSION | AY017378 | | |
| VERSION | AY017378.1 | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Homo sapiens. | | |
| REFERENCE | | | |
| AUTHORS | 1 (bases 1 to 3373) Coccone, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-J., Capron, A. and Bahr, G.M. | | |
| TITLE | Identification of a new RNA helicase (RH16) regulated by the Immunomodulator Murabudide | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 3373) Coccone, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-J., Capron, A. and Bahr, G.M. | | |
| AUTHORS | | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (09-JAN-2001) Laboratoire d'Immunologie Moléculaire de l'Infection et de l'Inflammation, Institut Pasteur de Lille, 1 rue du Professeur Calmette BP 245, Lille 59019, France | | |
| FEATURES | | | |
| source | 1..3373 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |

CDS 155..3232
/codon_start-1
/product-"RNA helicase-DEAD box protein #H16"
/db_xref-"GI:12621066"
/translation-"MSGNSTDENFRVYLISCPARVKNLIVPEVLYDTLELPAYRKE
OIOIRVATSGNMAVEELISTLEKGMVHMGWREVENFHTGSLARVKNPELTDL
PSPFENNADEYLOLNLLOPLVPLKLVRYVLDKCMHFTLTIDKRLIAEENGN
ESGVLELLKRVQENKMPFSAFLNLRVGNELVQLVPSKESNAFLENLSDVDK
OYEQLSTTVPLNEKEMVMSSESPDSVSVSNITSLAEVSGSLDESLGH
NSNMGSDSGTSGSDSEVAVARASPEPELQRPVOMELVAFKEPFLKNIICLPVGS
KTRVAVYIAKHLDDKKKASPEKVIYVANKVLEVDVAFKEPFLKNIICLPVGS
DTOLKISPEVYVSCDIIISTAOILENLIENMEDDAVLSDSFLIIDGHTTNK
EAVYNNINRHYLMOKLKNRKLKRNKRPILPDIIGLTNFGVGAIKQAKAEHLK
LCANDALFTIKVKNELDLKNOIQEPKFAIADTRHPEKLELNTLRIQTYCO
MSPSDFQPYEOMAIOMEKKAEGNKESEVAKYNNKALQINDIIMIDAVT
HLETFYNEEKDKFAVIEDSDGDEDEDEDEDKKPLKLEIDRELFMTLEFE
NNMKILRLAENPEYENELKTLRNTIMQYTRTEESARLITTRVAVYASLQMTIE
NEKFAVCYKAHLIGAGHSEFKPTONFOKEVSKSKINILNLTATVAEELDI
KECNVIRIGLVLTNFIAYVOAGRARADESTVLAHNSVITREIVNDEKEMVYK
ALHCVQNMKPEEYARHLELOMOSIMEKIKTRIAKYNKNSPLIFLCNCSYLA
CSGEDIHVIENKHNHNTPEPEKELYIVENKALQKCAVYINGEILCKGAGMTTM
VHKGLDPLKLRNFVVVVKNNSTKKYKKNVELPITFLNIDYSCILPDED"
1148..1201
/misc_feature
/note-"Region: DEAD-like helicase domain"
1481..1627
/feature
/region: DEAD/DEAH box"
/misc_feature
/feature
/region: conserved C-terminal helicase region: Region:
helicase C"
BASE COUNT 1158 a 637 c 748 g 830 t
ORIGIN

Query Match 99.2%; Score 3338.8; DB 9; Len: 113373;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3343; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

15 GACCCCTGTGGACAACTGCTATTTGTGAGGACAGACGGGTGATGACCCCTTTCTTAAAG 74
1 GGGCCCTGTGGACAACTGCTATTTGTGAGGACAGACGGGTGATGACCCCTTTCTTAAAG 60
75 TGGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 134
61 TGGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 120
135 TGGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 194
121 TGGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 180
195 GAATTCGGTATCATCTCGTGTGAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 254
181 GAATTCGGTATCATCTCGTGTGAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 240
255 GCGCTGTGAGTACCTGACCTTTGCTGAGAGAGTGAAGAGACAGGACAGGACAGGACAGG 314
241 GCGCTGTGAGTACCTGACCTTTGCTGAGAGAGTGAAGAGACAGGACAGGACAGGACAGG 300
315 AGTGGCACCTCCGCGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 374
301 AGTGGCACCTCCGCGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 360
375 AGTGGCACCTCCGCGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 434
361 AGTGGCACCTCCGCGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGG 420
435 TCTGGCGCGCGGTACATGAAGCTGAGCTCAAGGACCTTCCCTCCACATTTTGTAGAA 494
421 TCTGGCGCGCGGTACATGAAGCTGAGCTCAAGGACCTTCCCTCCACATTTTGTAGAA 480
495 CGCTCATGATGATATCTCCAACTGCTGAAGCTCCTTCAGGACAGCTTAAGACAGCT 554
481 CGCTCATGATGATATCTCCAACTGCTGAAGCTCCTTCAGGACAGCTTAAGACAGCT 540

555 TCTAGTAGACAGCTCTTGGATGAATGATGAGAGGAGAACTGTTGACATTTGAAGTAG 614
541 TCTAGTAGACAGCTCTTGGATGAATGATGAGAGGAGAACTGTTGACATTTGAAGTAG 600
615 AAACCGGATTCGCTGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 674
601 AAACCGGATTCGCTGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 660
675 AAGGATTCGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 734
661 AAGGATTCGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 720
735 AAACATGATTCGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 794
721 AAACATGATTCGCGAGAAACATGAAATGAATCAGGTGAACAGATTAATAA 780
795 TAGAATTTATACAAAGTTGATGTCGTCGCAAGTGGACAGCAACTCTTCAACAGT 854
781 TAGAATTTATACAAAGTTGATGTCGTCGTCGCAAGTGGACAGCAACTCTTCAACAGT 840
855 TCAGCAATTCGCGAGAAAGAGGAGTGTGGGCGCATGGAGAAATCACTCATGAAATCTTT 914
841 TCAGCAATTCGCGAGAAAGAGGAGTGTGGGCGCATGGAGAAATCACTCATGAAATCTTT 900
915 TCGACATTCCTGCTGATGTTTGAATTCAGACACAAAGTTTGGCAGAGCAAGTCTACTG 974
901 TCGACATTCCTGCTGATGTTTGAATTCAGACACAAAGTTTGGCAGAGCAAGTCTACTG 960
975 CTTAGATGAAAGTCTTGGACATTAACAGCAACATGGGAGTATTAAGCAATTAAGCAAG 1034
961 CTTAGATGAAAGTCTTGGACATTAACAGCAACATGGGAGTATTAAGCAATTAAGCAAG 1020
1035 TGATTCAGATGAAGAAATGTGGACAGCAAGACATCCCGGACAGCAAGTCTTAAGCAAG 1094
1021 TGATTCAGATGAAGAAATGTGGACAGCAAGACATCCCGGACAGCAAGTCTTAAGCAAG 1080
1095 GCGTTACCAAAATGGAAGTTGCCAGCAGCAGCTTGAAGAGGAAGATATCATCATTTACTG 1154
1081 GCGTTACCAAAATGGAAGTTGCCAGCAGCAGCTTGAAGAGGAAGATATCATCATTTACTG 1140
1155 CCGTACAGGAGTGGAAAAACAGAGTGGCTGTATTCAGTCAAGATCACTTAAGCAAG 1214
1141 CCGTACAGGAGTGGAAAAACAGAGTGGCTGTATTCAGTCAAGATCACTTAAGCAAG 1200
1215 GAAGAAAAAGCATCTGAGCCTGGAAGAAATTAATCTTGTCAATAAGTATCTTAAGT 1274
1201 GAAGAAAAAGCATCTGAGCCTGGAAGAAATTAATCTTGTCAATAAGTATCTTAAGT 1260
1275 TGAACAGCTCTCGGCAAGGAGTTCACACATTTTGAAGAAAGTATGCTGTTATTGG 1334
1261 TGAACAGCTCTCGGCAAGGAGTTCACACATTTTGAAGAAAGTATGCTGTTATTGG 1320
1335 ATTAAGTGTGATACCACTGAAGAAATTAATCTTGTCAATAAGTATCTTAAGT 1394
1321 ATTAAGTGTGATACCACTGAAGAAATTAATCTTGTCAATAAGTATCTTAAGT 1380
1395 TATTATCAGTACAGTCAAAATCTTGAAGAACTCCCTTTAAACTTGAAGAAATGGAAGA 1454
1381 TATTATCAGTACAGTCAAAATCTTGAAGAACTCCCTTTAAACTTGAAGAAATGGAAGA 1440
1455 TCGCTGTGTCATTAATGAGAACTTTCCCTCATTAATCAATGATGAAATGTCATCAATCA 1514
1441 TCGCTGTGTCATTAATGAGAACTTTCCCTCATTAATCAATGATGAAATGTCATCAATCA 1500
1515 CAAGAACAGAGTATTAATTAATCATGATGAGCAATTAATTAATGAGAAATGGAAGA 1574
1501 CAAGAACAGAGTATTAATTAATCATGATGAGCAATTAATTAATGAGAAATGGAAGA 1560
1575 TAGACTCAAGAAAGAAACCAAGATGATTCCTTCAGATGAGGAGCTTAAGAGC 1634
1561 TAGACTCAAGAAAGAAACCAAGATGATTCCTTCAGATGAGGAGCTTAAGAGC 1620
1635 TTGACCTGTGTTGAGAGGCGCCAGCAAGCAAGCAAGCTGAAGACAGATTTTAAGACT 1694

BASE COUNT 666 a 291 c 363 g 456 t
 ORIGIN
 Query Match 52.5%; Score 1766; DH 9; Length 1776;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1769; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1592 ACAACCAAGTATCCCTCTCTGAGATGAGGACTAACAGCTTCACCTGCTGTTGAG 1651
 2 ACAACCAAGTATCCCTCTCTGAGATGAGGACTAACAGCTTCACCTGCTGTTGAG 61
 1652 GGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1711
 62 GGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
 1712 GATTACTATTAACCTGTTAAAGAAACCTGATCAAGGAAACCAATACAGAGAG 1771
 122 GATTACTATTAACCTGTTAAAGAAACCTGATCAAGGAAACCAATACAGAGAG 181
 1772 CATGCAAGAGTTGGCCATTCAGATGACACAGAGAGAGATCCATTAAACAGAAC 1831
 182 CATGCAAGAGTTGGCCATTCAGATGACACAGAGAGAGATCCATTAAACAGAAC 241
 1832 TACAATTAATGACAGAGATTCACCTATTTGCAATGAGTCCAAATGTTGGA 1891
 242 TACAATTAATGACAGAGATTCACCTATTTGCAATGAGTCCAAATGTTGGA 301
 1892 CTCACCCCTATGACAGAGATTCACCTATTTGCAATGAGTCCAAATGTTGGA 1951
 302 CTCACCCCTATGACAGAGATTCACCTATTTGCAATGAGTCCAAATGTTGGA 361
 1952 GCAAAAGAGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
 362 GCAAAAGAGTGTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
 2012 ACACATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2071
 422 ACACATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
 2072 ATAGAAGTTGGAGTATGAGAGATGATGATGATGATGATGATGATGATGATG 2131
 482 ATAGAAGTTGGAGTATGAGAGATGATGATGATGATGATGATGATGATGATG 541
 2132 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2191
 542 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 601
 2192 TTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2251
 602 TTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 661
 2252 AATATGAAATGAAAGTGCAGCAATTAAGAAATGAAATGAAATGAAATGAAAT 2311
 662 AATATGAAATGAAAGTGCAGCAATTAAGAAATGAAATGAAATGAAATGAAAT 721
 2312 CTGAGAGATGACAGAGAGATTAATCTTACAAAGAGAGAGAGAGAGAGAGAG 2371
 722 CTGAGAGATGACAGAGAGATTAATCTTACAAAGAGAGAGAGAGAGAGAGAG 781
 2372 CCCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2431
 782 CCCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 841
 2432 TTGAGAGTGCAG 2491
 842 TTGAGAGTGCAG 901
 2492 TTAGTAATTTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2551
 902 TTAGTAATTTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
 2552 GTCTGATATTAAGAGATTAAGATGATGATGATGATGATGATGATGATGATG 2611

Db 962 GTCTGATATTAAGAGATTAAGATGATGATGATGATGATGATGATGATGATG 1021
 QY 2612 CCATGTCAG 2671
 Db 1022 CCATGTCAG 1081
 QY 2672 ACAGTGTTCAGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2731
 Db 1082 ACAGTGTTCAGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
 QY 2732 ATAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2791
 Db 1142 ATAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1201
 QY 2792 AATTACAGATGCAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2851
 Db 1202 AATTACAGATGCAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
 QY 2852 ATTACAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2911
 Db 1262 ATTACAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
 QY 2912 GTTCTGGGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2971
 Db 1322 GTTCTGGGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1381
 QY 2972 TCAAGAGACTTACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3031
 Db 1382 TCAAGAGACTTACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441
 QY 3032 AATTAATGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3091
 Db 1442 AATTAATGAGAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1501
 QY 3092 AAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3151
 Db 1502 AAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1561
 QY 3152 CACAAAG 3211
 Db 1562 CACAAAG 1621
 QY 3212 ATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3271
 Db 1622 ATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1681
 QY 3272 TACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3331
 Db 1682 TACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1741
 QY 3332 ACATAGAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 3391
 Db 1742 ACATAGAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1801

RESULT 7
 AA300841 1443 bp DNA linear PAT 30-NOV-2001
 LOCUS
 DEFINITION Sequence 10 from Patent WO0185955.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 Bahr, G., Cocude, C. and Capron, A.
 Rh116 polypeptides and its fragments and polynucleotides encoding
 said polypeptides and therapeutic uses
 Patent: WO 0185955-A 10 15-NOV-2001;
 JOURNAL
 Istec (FR); INSTITUT PASTEUR DE LILLE (FR)
 FEATURES
 Location/Qualifiers

source 1.1443
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 508 a 261 c 313 g 360 t 1 oth: 15
 ORIGIN

Query Match 42.7%; Score 1435.5; DB 6; Length 1443;
 Best Local Similarity 99.7%; Pred. No. 1.1e-292;
 Matches 1438; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 687 GAAAGAAAGTGTCTCTGCAATTTTGAATGTTCTTCGCAACAGAAATTAATCACT 746
 DB 1 GAAAGAAAGTGTCTCTGCAATTTTGAATGTTCTTCGCAACAGAAATTAATCACT 60
 QY 747 TGTCCAGAGTTAAGAGCTGTGATGTCAGAAAGCATGCAAGATTTCAATTTATC 806
 DB 61 TGTCCAGAGTTAAGAGCTGTGATGTCAGAAAGCATGCAAGATTTCAATTTATC 120
 QY 807 ACAAGTTGATGCTCTCAAGTGGAGAGCAACTTTCTTCAACAGATTTCACAAATCT 866
 DB 121 ACAAGTTGATGCTCTCAAGTGGAGAGCAACTTTCTTCAACAGATTTCACAAATCT 180
 QY 867 GGAAGAGAGGTGTGGGGCATGAGAAATTAATCATAGATCATTTTTCAGATTTCTG 926
 DB 181 GGAAGAGAGGTGTGGGGCATGAGAAATTAATCATAGATCATTTTTCAGATTTCTG 240
 QY 927 TGTAGTTTGAATGACAGCAAGTTTGGCAGAGAGATGTGCTGATGATGAAG 986
 DB 241 TGTAGTTTGAATGACAGCAAGTTTGGCAGAGAGATGTGCTGATGATGAAG 300
 QY 987 TCTTGACATTAACAGACATGGGCGATGATTCAGGACCATGGAAGTTCACAGATCA 1046
 DB 301 TCTTGACATTAACAGACATGGGCGATGATTCAGGACCATGGAAGTTCACAGATCA 360
 QY 1047 AGAGATTTGGCGAGCAAGACATCCCGGAGCCAGAACTTCAGCTAGGCTTACCAAT 1106
 DB 361 AGAGATTTGGCGAGCAAGACATCCCGGAGCCAGAACTTCAGCTAGGCTTACCAAT 420
 QY 1107 GGAAGTTGCCCGCAGCCTTGGAGGGAAGATATCAATCTGCTCTGACAGGAG 1166
 DB 421 GGAAGTTGCCCGCAGCCTTGGAGGGAAGATATCAATCTGCTCTGACAGGAG 480
 QY 1167 TGGAAAACCGAGAGTGGCTTTTACATTTGCCAAGATCACTTACAGAGTAAAAAGC 1226
 DB 481 TGGAAAACCGAGAGTGGCTTTTACATTTGCCAAGATCACTTACAGAGTAAAAAGC 540
 QY 1227 ATCTGAGCCTGGAAAAAGTTATAGTCTTGTCAATTAAGGTAAGCTGCTTACAGCTCTT 1286
 DB 541 ATCTGAGCCTGGAAAAAGTTATAGTCTTGTCAATTAAGGTAAGCTGCTTACAGCTCTT 600
 QY 1287 CCGCAAGAGATTCACACCATTTTGAAGAAATGTATCGTGTATTTGATTAAGTGTGA 1346
 DB 601 CCGCAAGAGATTCACACCATTTTGAAGAAATGTATCGTGTATTTGATTAAGTGTGA 660
 QY 1347 TACCAACTGAAATATCATTTTCCAGAGTGTCAAGTCTGTATATTATTCAGTAC 1406
 DB 661 TACCAACTGAAATATCATTTTCCAGAGTGTCAAGTCTGTATATTATTCAGTAC 720
 QY 1407 AGCTCAATCTTGAAGAACTCCTCTTAACTTGAAGAAATGAGAGATTTGCTGTCA 1466
 DB 721 AGCTCAATCTTGAAGAACTCCTCTTAACTTGAAGAAATGAGAGATTTGCTGTCA 780
 QY 1467 ATGTGAGATTTTCCCTCATTTATCATGTAGTAAGTGCATACAGCAAGTACAGCACT 1526
 DB 781 ATGTGAGATTTTCCCTCATTTATCATGTAGTAAGTGCATACAGCAAGTACAGCACT 840
 QY 1527 GATTAATTAACATCATGAGCATTAATTTGATGAGAGTTGAAAAACATTAATCAAGAA 1586
 DB 841 GATTAATTAACATCATGAGCATTAATTTGATGAGAGTTGAAAAACATTAATCAAGAA 900
 QY 1587 AGAAAACAAACGAGTATTCCTCTCTCAGATAGTGGAGTAAAGCTTACCTGTGTG 1646
 DB 901 AGAAAACAAACGAGTATTCCTCTCTCAGATAGTGGAGTAAAGCTTACCTGTGTG 960

QY 1647 TGGAGGGGGCCAGAGCAAGCAAGCTGAAGACATTTTAAACTATGTC AATCT 1706
 DB 961 TGGAGGGGGCCAGAGCAAGCAAGCTGAAGACATTTTAAACTATGTC AATCT 1020
 QY 1707 TGATGATTTACTATTAAACTGTTAAGAAAAACCTTGATCACTGAAACCAATACA 1766
 DB 1021 TGATGATTTACTATTAAACTGTTAAGAAAAACCTTGATCACTGAAACCAATACA 1080
 QY 1767 GGAGCCATGCAAGAGTTTCCCATTTGCAAGATGCAACGAGAGATTCATTTAAAGCA 1826
 DB 1081 GGAGCCATGCAAGAGTTTCCCATTTGCAAGATGCAACGAGAGATTCATTTAAAGCA 1140
 QY 1827 ACTCTAGAAATTAATGACAGAGATTCAAACTATTGTCAATAGATTCATGATTT 1886
 DB 1141 ACTCTAGAAATTAATGACAGAGATTCAAACTATTGTCAATAGATTCATGATTT 1200
 QY 1887 TGGAACTCAACCTTATGAAACATGGGGCATTCCAATTAAGAAAAAGCTGCAAAAAAG 1946
 DB 1201 TGGAACTCAACCTTATGAAACATGGGGCATTCCAATTAAGAAAAAGCTGCAAAAAAG 1260
 QY 1947 AATGCGAAGAGAGTGTGTTGCGAAGACATTTGAGAGTCAATGAGCCCTACAAAT 2006
 DB 1261 AATGCGAAGAGAGTGTGTTGCGAAGACATTTGAGAGTCAATGAGCCCTACAAAT 1320
 QY 2007 TAATGACAAATTTGAAATGATGATGCGTATGCTCATGCTTGAACCTTCTAATGAAGA 2066
 DB 1321 TAATGACAAATTTGAAATGATGATGCGTATGCTCATGCTTGAACCTTCTAATGAAGA 1380
 QY 2067 GAAAGATTAAGAGTTTGGCATGATGAGATGATGATGAGAGGTGTGATGATGATCA 2126
 DB 1381 GAAAGATTAAGAGTTTGGCATGATGAGATGATGATGAGAGGTGTGATGATGATCA 1440
 QY 2127 TTG 2129
 DB 1441 TTG 1443

RESULT 8
 LOCUS BC025508 2468 bp mRNA linear 07-AUG-2002
 DEFINITION Mus musculus, clone MGC:38064 IMAGE:5252257, mRNA, complete cds.
 ACCESSION BC025508
 VERSION BC025508.1 GI:19343862
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2468)
 Strausberg, R.
 Direct Submission
 Submitted (06-MAR-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gagbs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (NIH)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Galthersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akher, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, P., Brooks, S.,
 Diatcheva, N.L., Granite, S., Guan, X., Gupta, D., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Gasp, R.,
 Maduro, O.L., Mastello, C., Maskell, B., Mastrian, S.D., McCluskey, J.C.,
 McDowell, J., Pearson, R., Stantrilop, S., Thomas, P., Touchman, J.W.,

| | | | |
|----------------------------|---|---|-----------------------|
| Dd | 1564 | CACCGAATTCGAGGAGCTCTTCAATTGTGAAGAAAAACAAGACTGCATAAGAATTGG | 1623 |
| Oy | 3023 | CCGACTATCAATAAATGTTGGTAATCATCTGCCAAATGTGGCCAGGCTTGCCAACATGA | 3082 |
| Dd | 1624 | CTGATTATCAGACCMAATGAGAGATTATCTGCAGAGTGTGGCAGGCTTGCCAACATGA | 1683 |
| Oy | 3083 | TGGTCACAAAAGCGCTACATTTTGGCTTGTCTCAAAATAGCATTTTGTATGGTTTTCA | 3142 |
| Dd | 1684 | TGGTCACAAAAGCGTTACATTTTGGCTTGTCTTAAATAGCATTTTGTATGGTTTTCA | 1743 |
| Oy | 3143 | AAAAATATCAACAAAGAAACATATCCAAAAGTGGTGTGAATTACTTAAATATTTCCCA | 3202 |
| Dd | 1744 | AAAAATATCTCACCGAAGAAACGTACAAAGAGTGGTGTGAATTGGCTTAATTTCTCG | 1803 |
| Oy | 3203 | AATCTGACTATTCAGAAATGCTGTTTATTTTAGTAGAGAGATTAGACTTAATGAAAT | 3262 |
| Dd | 1804 | ATCTGACTACTACGAATATAGCTGTTATAGTAGAAGAGATTAGACTTAATGAAAT | 1863 |
| Oy | 3263 | CTTTAAAAATACATACGTTAAACATTTAATATGATATGATTAATGTAATTAATGCT | 3322 |
| Dd | 1864 | ATTTAAAAATATGTCCAACTCAACATTAATATGATTTAATTTGTTTCTCTACACT | 1923 |
| Oy | 3323 | ACAGACGACATAGAAATCAATAAA | 3348 |
| Dd | 1924 | ACTGAGCTACTCTAGAAATATGTAGA | 1949 |
| RESULT | 9 | | |
| LOCUS | AX300838 | 1284 bp | DNA |
| DEFINITION | Sequence 7 from patent WO0185955. | | LINE: PAT 30-NOV-2001 |
| ACCESSION | AX300838 | | |
| VERSION | AX300838.1 | GI:17382116 | |
| KEYWORDS | human. | | |
| SOURCE | ORGANISM Homo sapiens | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| REFERENCE | 1 | | |
| AUTHORS | Bahr,G., Cocude,C. and Capron,A. | | |
| TITLE | Rhl6 polypeptides and its fragments and polynucleotides encoding said polypeptides and therapeutic uses | | |
| JOURNAL | Patent: WO 0185955-A 7 15-NOV-2001; | | |
| | Istac (FR) ; INSTITUT PASTEUR DE LILLE (FR) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..1284 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| BASE COUNT | 489 a 198 c 261 g 335 t | 1 others | |
| ORIGIN | | | |
| Query Match | 37.3%; Score 1254.5; DB: 6; Length 1284; | | |
| Best Local Similarity | 99.6%; Pred.No. 1.8e-254; | | |
| Matches 1257; Conservative | 0; Mismatches 5; Indels 0; Gaps 0. | | |
| Oy | 2103 | TGATAGGCTGCTGATGATGAGTATTGTGATGCTGATGAGATAGAGATATTAAAGAA | 2162 |
| Dd | 1 | TGATAGGCTGCTGATGATGAGTATTGTGATGCTGATGAGATAGAGATATTAAAGAA | 60 |
| Oy | 2163 | ACCTTTGAAACTGTAGTAAACAGATAGATTTCATGACTTTATTTTTTAAACATAA | 2222 |
| Dd | 61 | ACCTTTGAAACTGTAGTAAACAGATAGATTTCATGACTTTATTTTTTAAACATAA | 120 |
| Oy | 2223 | AATGTTGAAAAGGCTGGCTGAAAACCCAGCATATGAAAATGAAAAGCTGCTAAATTAAG | 2282 |
| Dd | 121 | AATGTTGAAAAGGCTGGCTGAAAACCCAGCATATGAAAATGAAAAGCTGCTAAATTAAG | 180 |
| Oy | 2283 | AAATACCATTAAGGAGCATATAGTAGAGCTGAGGATAGACAGATAGAAATATCTTTAC | 2342 |
| Dd | 181 | AAATACCATTAAGGAGCATATAGTAGAGCTGAGGATAGACAGATAGAAATATCTTTAC | 240 |
| Oy | 2343 | AAAAACACGACAGAGTATGATGCGTTTCCAGTAGTGATTACTTGAAAAATGAAATTTGC | 2402 |

| | | | |
|----|------|--|------|
| Db | 241 | AAAAACCGACAGATGTCATATGCGCTTTCCAGTGGATTTACTGAAAAATTGCAAAATTGCG | 300 |
| QY | 2403 | TGAAGTGGAGATCAAAAGCCACCATTCTGATTTGGAGCTGGACACACAGCTGAGTTAAAC | 2462 |
| Db | 301 | TGAAGTGGAGATCAAAAGCCACCATTCTGATTTGGAGCTGGACACACAGCTGAGTTAAAC | 360 |
| QY | 2463 | CATGACACAGATGAAACAAAAAGAGTCAATGTAATTTGGCACTGGAAAAATTAATCT | 2522 |
| Db | 361 | CATGACACAGATGAAACAAAAAGAGTCAATGTAATTTGGCACTGGAAAAATTAATCT | 420 |
| QY | 2523 | GCTTATGCTACACAGTGGCCAAAGAGTGGATTTAAGAAATTAACATGTTAT | 2582 |
| Db | 421 | GCTTATGCTACACAGTGGCCAAAGAGTGGATTTAAGAAATTAACATGTTAT | 480 |
| QY | 2583 | CCGTTATGCTGCTGTCACCAATGAAATPAGCCATGTGTCAGGCCGTGGTCAGGCATAGC | 2642 |
| Db | 481 | CCGTTATGCTGCTGTCACCAATGAAATPAGCCATGTGTCAGGCCGTGGTCAGGCATAGC | 540 |
| QY | 2643 | TGATGAGAGCACCAGTCTGCTGGTTGCTCACAGTGGTTGAGAGTTATGCAACATAGAC | 2702 |
| Db | 541 | TGATGAGAGCACCAGTCTGCTGGTTGCTCACAGTGGTTGAGAGTTATGCAACATAGAC | 600 |
| QY | 2703 | AGTTAATGATTTCCGAGAGAAAGTATGTTAAGCTATACATTGTTGTTCAAAATATGAA | 2762 |
| Db | 601 | AGTTAATGATTTCCGAGAGAAAGTATGTTAAGCTATACATTGTTGTTCAAAATATGAA | 660 |
| QY | 2763 | ACCGAGAGAGTATGCTCATAGATTTTGGATTACAGATGCAAGATTAATGAGAAAGAA | 2822 |
| Db | 661 | ACCGAGAGAGTATGCTCATAGATTTTGGATTACAGATGCAAGATTAATGAGAAAGAA | 720 |
| QY | 2823 | AATGAAACCAAGAGAAATATTGCCAAGCATTAACAAGATTAACCCATCACTATATACATT | 2882 |
| Db | 721 | AATGAAACCAAGAGAAATATTGCCAAGCATTAACAAGATTAACCCATCACTATATACATT | 780 |
| QY | 2883 | CCTTTGCAAAAAATGCGAGTGTGCTAGCCGTGTTCTGGGGAAAGTATCCATCTATATGACAA | 2942 |
| Db | 781 | CCTTTGCAAAAAATGCGAGTGTGCTAGCCGTGTTCTGGGGAAAGTATCCATCTATATGACAA | 840 |
| QY | 2943 | AATGATCAGCTGCAATATGACCCCAATTCAGAAAGAACTTTACATTGTAAGAGATTAACA | 3002 |
| Db | 841 | AATGATCAGCTGCAATATGACCCCAATTCAGAAAGAACTTTACATTGTAAGAGATTAACA | 900 |
| QY | 3003 | AGCACTGCAAAAAGAGTGTCCGACACTATCAAAATAAATGTTGAAATCATCTCAAAATGCG | 3062 |
| Db | 901 | AGCACTGCAAAAAGAGTGTCCGACACTATCAAAATAAATGTTGAAATCATCTCAAAATGCG | 960 |
| QY | 3063 | CCAGGCTTTGGGGAACAATGATGTTGTCACAAAGGCTTAGATTTCCTGTCTAAATTAAG | 3122 |
| Db | 961 | CCAGGCTTTGGGGAACAATGATGTTGTCACAAAGGCTTAGATTTCCTGTCTAAATTAAG | 1020 |
| QY | 3123 | GAATTTGTGATGTTTTGAAAAATTAATCAACAAAGAAACATACAAAAAGTGGTGA | 3182 |
| Db | 1021 | GAATTTGTGATGTTTTGAAAAATTAATCAACAAAGAAACATACAAAAAGTGGTGA | 1080 |
| QY | 3183 | ATTACCTATCACATTTCCCATCTTTCAGTATTCAGATGCTGTTATTATAGTCAACAGAA | 3242 |
| Db | 1081 | ATTACCTATCACATTTCCCATCTTTCAGTATTCAGATGCTGTTATTATAGTCAACAGAA | 1140 |
| QY | 3243 | TTAGCACTGATTTGAGAGATTTTAAATTAATCAATGTTAAACATTTAATATATG | 3302 |
| Db | 1141 | TTAGCACTGATTTGAGAGATTTTAAATTAATCAATGTTAAACATTTAATATATG | 1200 |
| QY | 3303 | ATTATATGATTCATTATGCTACAGAACTGACATTAACAATCAATTAATATATGTTTACT | 3362 |
| Db | 1201 | ATTATATGATTCATTATGCTACAGAACTGACATTAACAATCAATTAATATATGTTTACT | 1260 |
| QY | 3363 | CT 3364 | |
| Db | 1261 | NT 1262 | |

Db 1138 AGCATTCGAAAGCAATACAGCAATCCATCGTTAATACATTTCTTCTTAAAAATTGT 1197
 QY 2899 AGTGTCTAGCCCTGTTCTGTGGGAGAGATATCCATGTAATAGAAAAATGTTACGTCAT 2958
 Db 1198 AGCATTCGTTCTGTGGGAGAGAAACATTCATCGTATCGAGAAAGATGTTATGTCAT 1257
 QY 2959 ATGACCCCAATTCAGAACTTACATTTGTAAGAAAAACAACACATCTTAAAAAAG 3018
 Db 1258 ATGACACCAATTCAGAAAGCAATCTGTAAGAAAAACAACCAATCTTAAAAAAGAA 1317
 QY 3019 TGTGCCACTATCAATTAATGTTGTAATCATCTGCAAAATGTGGCCAGGTTGGGAGAA 3078
 Db 1318 TTGTGATTAATACAGCAATGAGAGATTAATCTGCAATGTGGCCAGGTTGGGAGAA 1377
 QY 3079 ATGATGTCACAAAGCGTTAGATTGCTGTCTGCAAAATAGAAATTTATGTCCTT 3138
 Db 1378 ATGATGTCACAAAGCGTTAGATTGCTGTCTGCAAAATAGAAATTTATGTCAT 1437
 QY 3139 TTCAAAAATTAATTCACAAAGCAATTCACAAATAGTGGGTAGATTAATATCATATT 3198
 Db 1438 TTCAAAAATTAATTCACAAAGCAATTCACAAAGCAATGCTGGAATGCTATCATATT 1497
 QY 3199 CCCAATCTGACTATTCAGAACTGCTTTATTAGTATGAGATGAGATGAGATGATTGA 3258
 Db 1498 CCGATCTTGTGACTACAGCAATGCTGCTGTATAGTATGAGATGAGATGAGATGATTGA 1557
 QY 3259 GATCTTTTAAATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3318
 Db 1558 GATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1617
 QY 3319 TGCTACAGCAATGACATTAAGATCAATTA 3348
 Db 1618 CACTACTGAGCTACTCTAGAAATATGTA 1647

RESULT 12
 AX098232 1258 bp DNA linear PAT 10-MAR-2001
 LOCUS Sequence 7 from Patent W00118208.
 DEFINITION AX098232
 ACCESSION AX098232.1 GI:13515364
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
 1 (bases 1 to 1258)
 AUTHORS Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
 TITLE Interferon induced polynucleotides and proteins encoded thereby
 JOURNAL Patent: WO 0118208-A 7, 15-MAR-2001;
 Curegen Corporation (US); BIOGEN, INC. (US)
 FEATURES
 source Location/Qualifiers
 1..1258
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..915
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC35440.1"
 /db_xref="GI:13515364"
 CDS
 1..1258
 /translation="MVNSNTTPRRDVIKSPVAGSRPLSKMAADSGTHISPYPTT
 APSAENPLPSLRKDDVFWFHRREFVLSHVLQCGGKAVPEVKEOIORVATSGNM
 QAVELLSTLGGVHLDGTRFEVFALEKRTSPLAARVNEELDLPSPFENADEY
 LQNLNLOPTLVLDKLDVADVADKMEELLITEDRNITAAENNGNESGVRELLKRIY
 OKENMFSAFLVRLGTNNELVQELTSGDGSFNAEINLISQVDPQVEQLLSTVQ
 PNLKEVKGWMEKNSSEFSFADSVVSGT"

BASE COUNT 399 a 267 c 276 g 316 t
 ORIGIN

Query Match 23.9%; Score 803; DB 6; Length 1258;
 Best Local Similarity 95.5%; Pred. No. 4,5e-159;
 Matches 853; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 45 GCACAGAGCGGTAGACCCCTGCTTCTTAAGTGGGACGGGACGCGGATATATCA 104
 Db 55 GCACAGAGCGGTAGACCCCTGCTTCTTAAGTGGGACGGGACGCGGATATATCA 114
 QY 105 CCGTCCCGGACAGCAAGACCCATCTCTGGGAAACCCCTCCCTTCCTTGAAG 164
 Db 115 CCGTCCCGGACAGCAAGACCCATCTCTGGGAAACCCCTCCCTTCCTTGAAG 174
 QY 165 AAGATGTCGAATGGGTATTCACAGAGCAATTTCCGCTATCTCATCTCTG 224
 Db 175 AAGATGTCGAATGGGTATTCACAGAGCAATTTCCGCTATCTCATCTCTG 234
 QY 225 GCGCAGGCGTAAATATCATCCAGCTGAGCGCTGTGTGACTACCTTCTTCC 284
 Db 235 GCGCAGGCGTAAATATCATCTCTGAGCGCTGTGTGACTACCTTCTTCC 296
 QY 295 TGCAAGGTGAAAGAGCAATTCAGAGCAAGTCCGACCTCCGGAGCATTTA 344
 Db 257 --CAGAGGTGAAAGAGCAATTCAGAGCAAGTCCGACCTCCGGAGCATTTA 354
 QY 345 TGAATGCTGTGACACCTTGGAGAGGAGTGTGACCTTGTGACCTGAGCAAT 404
 Db 315 TGAATGCTGTGACACCTTGGAGAGGAGTGTGACCTTGTGACCTGAGCAAT 374
 QY 405 CCGTGAAGCCCTCCGAGAGACCGGACCTCTGCGCCCGCTACATGAATCTGACCT 464
 Db 375 CCGTGAAGCCCTCCGAGAGACCGGACCTCTGCGCCCGCTACATGAATCTGACCT 434
 QY 465 CAGGACCTTCCCTCTGATGTTGAGAGCGCTATGTAATATCCCAATCTGAA 524
 Db 435 CAGGACCTTCCCTCTGATGTTGAGAGCGCTATGTAATATCCCAATCTGAA 494
 QY 525 CCGCTTGAAGCCCTCTGATGTTGAGAGCGCTATGTAATATCCCAATCTGAA 584
 Db 495 CCGCTTGAAGCCCTCTGATGTTGAGAGCGCTATGTAATATCCCAATCTGAA 554
 QY 585 GGAGAGGAGACCTGTGACAAATTTGAAGACAGAAACCGGATTTGCTGAGAAATATG 644
 Db 555 GGAGAGGAGACCTGTGACAAATTTGAAGACAGAAACCGGATTTGCTGAGAAATATG 614
 QY 645 AATGAATCAGGTGTAAGAGAGCTACTAAAAAGATTTGAGAGAAAGAAATGCTCTC 704
 Db 615 AATGAATCAGGTGTAAGAGAGCTACTAAAAAGATTTGAGAGAAAGAAATGCTCTC 674
 QY 705 TGCAATTTCTGAATGTTCTTCTGCAAAACAGAAACATGAATCTGTCAAGAATTAACAG 764
 Db 675 TGCAATTTCTGAATGTTCTTCTGCAAAACAGAAACATGAATCTGTCAAGAATTAACAG 734
 QY 765 CCGTATTTCTGCAAAACAGAAACATGAATCTGTCAAGAATTTATCAAGTGAATGCTCA 824
 Db 735 CCGTATTTCTGCAAAACAGAAACATGAATCTGTCAAGAATTTATCAAGTGAATGCTCA 794
 QY 825 AGTGAAGAGCAATCTTTTCAACGACAGTTCAGCAATCTGGAAGAGAAAGTATGAGG 884
 Db 795 AGTGAAGAGCAATCTTTTCAACGACAGTTCAGCAATCTGGAAGAGAAAGTATGAGG 854
 QY 885 CATGAGAAATTAATCATCAGAAATCTTTTGCAGATTTCTGTAGTTGAG 937
 Db 855 CATGAGAAATTAATCATCAGAAATCTTTTGCAGATTTCTGTAGTTGAG 907

RESULT 13
 AX098236 1270 bp DNA linear PAT 30-MAR-2001
 LOCUS Sequence 11 from Patent W00118208.
 DEFINITION AX098236
 ACCESSION AX098236.1 GI:13515371
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 1270)

AUTHORS Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
 TITLE Interferon induced polynucleotides and proteins encoded thereby
 JOURNAL Patent: WO 0118208-A 11-15-MAR-2001
 Curagen Corporation (US); BIOGEN, INC. (US)

FEATURES
 source Location/Qualifiers
 1..1270

CDs
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..927
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC35442.1"
 /db_xref="GI:13515372"
 /translation="MRAAMVNSNTPRRDVIKSPVAVGPPKESKMAADSGTHISPV
 POTAPSAMENPLSLRKDYEMVFRHREFPLSHLVLDYQENHPRFKEDIORVAT
 SGMOAVELLSTLEKGYWHGLMREFEALRRTGSPGASRYMNEFLDPSPEANA
 HDEYLOLNLPFLVDKLVLDVLDKMEBEELIIEENRNIAAENNGNSGYELL
 KRIYKEMFSAFLNVLROTGNNELVOLTSDPSESNAIENLSQYDGPVEEQLLS
 TVVPLEKEWGMENNSSESSFADSVVSGT"

BASE COUNT 401 a 271 c 279 g 319 t
 ORIGIN

Query Match 23.9%; Score 803; DB 6; length 1270;
 Best Local Similarity 95.5%; Pred. No. 4,5e-159;

Matches 853; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

45 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGGCGGAAACACGCAATTTCA 104
 67 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGGCGGAAACACGCAATTTCA 126
 105 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTTCTGCTGAGCAAG 164
 127 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTTCTGCTGAGCAAG 186
 165 AAGATGTGGAATGGGTATTCACAGAGAGAAATTTCCCTATCTGCTGCTGAG 224
 187 AAGATGTGGAATGGGTATTCACAGAGAGAAATTTCCCTATCTGCTGCTGAG 246
 225 GGCAGAGGGTAAATGTACATCCAGGTGGAGGCTGTGCTGAGACACCTGCTGCTGAG 284
 247 GGCAGAGGGTAAATGTACATCCAGGTGGAGGCTGTGCTGAGACACCTGCTGCTGAG 268
 285 TGCAGAGGTGAAGAGACATTCAGAGAGAGTGGCACTCTGAGAGAGAGAGAGAG 344
 269 --CAGAGGTGAAGAGACATTCAGAGAGAGTGGCACTCTGAGAGAGAGAGAGAG 326
 345 TGAAGTGTGCTGAGACCTTGGAGAGAGAGTGTGCTGAGAGAGAGAGAGAG 404
 327 TGAAGTGTGCTGAGACCTTGGAGAGAGAGTGTGCTGAGAGAGAGAGAGAG 386
 405 CGTGGAGGGCTCCGGAGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 464
 387 CGTGGAGGGCTCCGGAGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446
 465 CAGGAGCTTCCCT 524
 447 CAGGAGCTTCCCT 506
 525 CCT 584
 507 CCT 566
 585 GAG 644
 567 GAG 626
 645 AAGTGAATGAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
 627 AAGTGAATGAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 686
 705 TGAATTTGTAATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 764
 687 TGAATTTGTAATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 746

QY 765 CTCTGATTGCTCAGAAAGCAATGCAGAGATTGAGAAATTTATCAGAGTTGCTCTCA 824

DB 747 CTCTGATTGCTCAGAAAGCAATGCAGAGATTGAGAAATTTATCAGAGTTGCTCTCA 806

QY 825 ACTGAGAGAGCAACTTCTTTCACCCAGGTTGACCAATCTGAGAGAGAGAGTGTGGG 884

DB 807 AGTGAAGAGAGCAACTTCTTTCACCCAGGTTGACCAATCTGAGAGAGAGAGTGTGGG 866

QY 885 CATGAGAGAGTATCATCATGAGATCATCTTTTGGAGATTCTTGTAGTTTCA 937

DB 867 CATGAGAGAGTATCATCATGAGATCATCTTTTGGAGATTCTTGTAGTTTCA 919

RESULT 14

AX098234 3692 bp DNA linear PAT 30-MAR-2001

AX098234 Sequence 9 from Patent WO0118208.

AX098234.1 GI:13515368

KEYWORDS

ORGANISM human.

LOCUS AX098234 3692 bp DNA linear PAT 30-MAR-2001

DEFINITION Sequence 9 from Patent WO0118208.

ACCESSION AX098234

VERSION AX098234.1 GI:13515368

FEATURES

source

CDs

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..804
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC35441.1"
 /db_xref="GI:13515369"
 /translation="MVAASNTTPRRDVIKSPVAVGPPKESKMAADSGTHISPV
 APSAMENPLSLRKDYEMVFRHREFPLSHLVLDGSENVHPEVEKEDIORVATSGM
 QAVELLSTLEKGYWHGLMREFEALRRTGSPGASRYMNEFLDPSPEANAHEX
 LOLNLPFLVDKLVLDVLDKMEBEELIIEENRNIAAENNGNSGYELLKRIY
 OKEMMFSAFLNVLROTGNNELVOLTSDPSESNAIENLSQYDGPVEEQLLS
 TVVPLEKEWGMENNSSESSFADSVVSGT"

BASE COUNT 1166 a 709 c 806 g 1011 t

Query Match 19.7%; Score 663; DB 6; length 3692;
 Best Local Similarity 92.9%; Pred. No. 1.7e-129;

Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;

45 GCACAGAGGGGTAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 104
 55 GCACAGAGGGGTAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 114
 105 CCTGTCCCGCAGACACACCATCTGCTGGAGAACCTTCTGCTGAGCAAG 164
 115 CCTGTCCCGCAGACACACCATCTGCTGGAGAACCTTCTGCTGAGCAAG 174
 165 AAGATGTGGAATGGGTATTCACAGAGAGAAATTTCCCTATCTGCTGCTGAG 224
 175 AAGATGTGGAATGGGTATTCACAGAGAGAAATTTCCCTATCTGCTGCTGAG 234
 225 GGCAGAGGGTAAATGTACATCCAGGTGGAGGCTGTGCTGAGACACCTTCTGCTGAG 284
 235 GGCAGAGGGTAAATGTACATCCAGGTGGAGGCTGTGCTGAGACACCTTCTGCTGAG 256
 285 TGCAGAGGTGAAGAGACATTCAGAGAGAGTGGCACTCTGAGAGAGAGAGAGAG 344
 257 --CAGAGGTGAAGAGACATTCAGAGAGAGTGGCACTCTGAGAGAGAGAGAGAG 314
 345 TGAAGTGTGCTGAGACCTTGGAGAGAGAGTGTGCTGAGAGAGAGAGAGAG 404

Db 315 TGAACCTGCTGAGACCTTGGACAGGAGAGTCTGGACACCTTTGGTTGATTCGGGAAT 374
 Oy 405 CGTGAAGGCGCTCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATATATCTGAT 464
 Db 375 CGTGAAGGCGCTCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATATATCTGAT 434
 Oy 465 CAGGAGCTGGCGCTCTGCAATCTTTGAGAACGCTATATGAAATCTCTGACTGTGA 524
 Db 435 CAGGAGCTGGCGCTCTGCAATCTTTGAGAACGCTATATGAAATCTCTGACTGTGA 494
 Oy 525 CCTCTTCAGCCCTCTGCTGAGACAAAGCTCTAGTTAGAGAGCTTGTATAGAGTAT 584
 Db 495 CCGCTTCAGCCCTCTGCTGAGACAAAGCTCTAGTTAGAGAGCTTGTATAGAGTAT 554
 Oy 585 GAGAGGAGAACTGTTGACAAATGGAACAGAAACCGGATGCTGTGCAATAACAGATG 644
 Db 555 GAGAGGAGAACTGTTGACAAATGGAACAGAAACCGGATGCTGTGCAATAACAGATG 614
 Oy 645 AATGAATGAGTGTAAAGAGAGTACTATAAAGGATTTGGCAAGAAATGCTGTC 704
 Db 615 AATGAATGAGTGTAAAGAGAGTACTATAAAGGATTTGGCAAGAAATGCTGTC 674
 Oy 705 TGCATTCTGAATGCTCTGCTCAACAGAGAAACATGAGTTGCCAATAGTTAAGAG 764
 Db 675 TGCATTCTGAATGCTCTGCTCAACAGAGAAACATGAGTTGCCAATAGTTAAGAG 734
 Oy 765 CTCTGATGCTCAGAAAGATGACAGATGAGAAATGATCAATAAGTTATGTC 821
 Db 735 CTCTGATGCTCAGAAAGATGACAGATGAGAAATGATCAATAAGTTATGTC 791

RESULT 15
 AX098238 3704 bp DNA 11n3 PAT 30-MAR-2001
 LOCUS AX098238
 DEFINITION Sequence 13 from Patent WO0118208.
 ACCESSION AX098238
 VERSION AX098238.1 GI:13515375
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 3704)
 PAYMAN, J. A., DA SILVA, A., HOCHMAN, P. and HAU, A.
 Interferon induced polynucleotides and proteins encoded thereby
 Patent: WO 0118208-A 13 15-MAR-2001;
 Curegen Corporation (US); BIOGEN, INC. (US)
 FEATURES
 source Location/Qualifiers
 1..3704
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..816
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC35443.1"
 /db_xref="GI:13515375"
 /translation="MPAAVMSNTTPRRDVIKSPVAV-GRPCFSKWAADSGTHISPV
 PQTAPSAMENPLPSLRKDYEVWFHREPFSLHLYD-GENVHPEVKEQIQRTVAT
 SGNQAVELLSTLEKGYHLGWTREPEALRRTGS-IAKYMNPETLDELSPSENA
 HDEYLILNLOPTLVDKLIVRDVLDKMEFEILLTE-ENRILAAENNGNSGVRELL
 KRIYKEMHFSFNLVLRQTNELNELYELTSDCESNAGICNTERRDSNSA"
 BASE COUNT 1168 a 713 c 809 g 1014 t
 ORIGIN

Query Match 19.7% Score 663; DB 6; Length 704;
 Best Local Similarity 92.9% Pred No. 1.7e-129;
 Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;
 Oy 45 GCACAGAGGGGTAGACCTGCTTCTTAAGTGGGAGCCAGAGCGGCAATTCATTTCA 104
 Db 67 GCACAGAGGGGTAGACCTGCTTCTTAAGTGGGAGCGAGAGCGGCAATTCATTTCA 126
 Oy 105 CCTGTCCCGCAGACACACCATCTGCTTGGGAGAGCTCTCTCCTTCTTGAGAAAG 164

Db 127 CCTGTCCCGCAGACACACCATCTGCTTGGGAGAAACCTCTCCTTCTTGAAG 186
 Oy 165 AAAGATGTGGAATGGATTTTCACAGAGAGAAATTTGCTATGCTATGCTGCTGAC 224
 Db 187 AAAGATGTGGAATGGATTTTCACAGAGAGAAATTTGCTATGCTATGCTGCTGAC 246
 Oy 225 GGCAGAGTGAATAATGATACAGAGTGGAGAGCTGTGCTGAGTACCTGACCTTTCTGCC 284
 Db 247 GGCAGAGTGAATAATGATACATC----- 268
 Oy 285 TGCAGAGTGAAGAGAGCATTCAGAGAGAGTGGCCACCTCCGGGAACATGACAGAGT 344
 Db 269 --CAGAGTGAAGAGAGCATTCAGAGAGAGTGGCCACCTCCGGGAACATGACAGAGT 326
 Oy 345 TGAACCTGCTGAGACACCTTGGAGAAAGAGTGTGACACCTTGGTGGAGCTCGGAAT 404
 Db 327 TGAACCTGCTGAGACACCTTGGAGAAAGAGTGTGACACCTTGGTGGAGCTCGGAAT 386
 Oy 405 CGTGAAGGCGCTCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATGAACCT 464
 Db 387 CGTGAAGGCGCTCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATGAACCT 446
 Oy 465 CAGGAGCTGGCCCTCTGCTGAGAGAGCTGATGATGATATATGCAACTGTGA 524
 Db 447 CAGGAGCTGGCCCTCTGCTGAGAGAGCTGATGATGATATATGCAACTGTGA 506
 Oy 525 CCTCTTCAGCCCTCTGCTGAGACAAAGCTTCTAGTTAGAGAGCTTGTGA1AA 584
 Db 507 CCTCTTCAGCCCTCTGCTGAGACAAAGCTTCTAGTTAGAGAGCTTGTGA1AA 566
 Oy 585 GAGAGGAGAACTGTTGACAAATGGAAGACAGAAACCGGATGCTGCTGACAGAA 644
 Db 567 GAGAGGAGAACTGTTGACAAATGGAAGACAGAAACCGGATGCTGCTGACAGAA 626
 Oy 645 AATGAATGAGTGTAAAGAGAGTACTATAAAGGATTTGGCAAGAAATGCTGTC 704
 Db 627 AATGAATGAGTGTAAAGAGAGTACTATAAAGGATTTGGCAAGAAATGCTGTC 686
 Oy 705 TGCATTCTGAATGCTCTGCTCAACAGAGAAACATGAGTTGCCAAGTTAAGAG 764
 Db 687 TGCATTCTGAATGCTCTGCTCAACAGAGAAACATGAGTTGCCAAGTTAAGAG 746
 Oy 765 CTCTGATGCTCAGAAAGATGACAGATGAGAAATGATCAATAAGTTATGTC 821
 Db 747 CTCTGATGCTCAGAAAGATGACAGATGAGAAATGATCAATAAGTTATGTC 803

Search completed: May 16, 2003, 18:23:30
 Job time : 8695 secs


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5786526"
/clone_id="NH_MGC_71"
/issue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sp1K16; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      383 a      161 c      229 g      240 t
ORIGIN
Query Match      21.2%; Score 714; DB 14; Length 1013;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 764; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1752 GAAAAACCAATACAGAGCCATGCAGAGTTGCCATTGCATGACACACAGAGA 1811
DB 1 GAAAAACCAATACAGAGCCATGCAGAGTTGCCATTGCATGACACACAGAGA 60
OY 1812 TCCATTAAAGAGAACTTCTAGAAATATGACAAAGATTCAACTTATGCTCAATGAG 1871
DB 61 TCCATTAAAGAGAACTTCTAGAAATATGACAAAGATTCAACTTATGCTCAATGAG 120
OY 1872 TCCATTGTCAGATTGTCAGTCAACCTATGACAAAGTGGCCATTCAATGSAAAAAA 1931
DB 121 TCCATTGTCAGATTGTCAGTCAACCTATGACAAAGTGGCCATTCAATGSAAAAAA 180
OY 1932 AGCTGCAAAAAAGAAATCGCAAGAAAGTGTGTCAGACAAATTTGAGGAAGTACAA 1991
DB 181 AGCTGCAAAAAAGAAATCGCAAGAAAGTGTGTCAGACAAATTTGAGGAAGTACAA 240
OY 1992 TGAGGCCCCACAAATATATGACAACTTGCATATGATGCTGATCTGCTTGAAC 2051
DB 241 TGAGGCCCCACAAATATATGACAACTTGCATATGATGCTGATCTGCTTGAAC 300
OY 2052 TTTCTATATGAGAGAAAGATTAAGAAAGTTGACATGACAAATGATGATGAGAG 2111
DB 301 TTTCTATATGAGAGAAAGATTAAGAAAGTTGACATGACAAATGATGATGAGAG 360
OY 2112 TGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2171
DB 361 TGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
OY 2172 ACTGATGAAACAGATATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2231
DB 421 ACTGATGAAACAGATATCTGATGATGATGATGATGATGATGATGATGATGATGAT 480
OY 2232 AAGCTGGCTGCAAAACCCAGATATGAAATGAAAGCTGAAATTAACAAATACCAT 2291
DB 481 AAGCTGGCTGCAAAACCCAGATATGAAATGAAAGCTGAAATTAACAAATACCAT 540
OY 2292 AATGAGCAATATCTAGAGCTGAGGAATGACAGACAGCAATATCTTAAATAACAC 2351
DB 541 AATGAGCAATATCTAGAGCTGAGGAATGACAGACAGCAATATCTTAAATAACAC 600
OY 2352 ACAGAGTCATATGCGCTTCCGAGTGTAGTGAATTAATAATTTGTAGTAGG 2411
DB 601 ACAGAGTCATATGCGCTTCCGAGTGTAGTGAATTAATAATTTGTAGTAGG 660
OY 2412 AGTCAAAAGCCAGCATCTGATTTGAGAGTGACAGACAGAGTGTGCAAAAGCTATGACA 2471
DB 661 AGTCAAAAGCCAGCATCTGATTTGAGAGTGACAGACAGAGTGTGCAAAAGCTATGACA 720
OY 2472 GAATGACAAAAAGAGTATTAGTAAATTTGCACTGCGAAAAAT 2516
DB 721 GAATGACAAAAAGAGTATTAGTAAATTTGCACTGCGAAAAAT 765

RESULT 2
BO772836/c      755 bp      mRNA      linear      EST 26-JUL-2002
LOCUS      BO772836
DEFINITION      UI-H-FEO-bdn-c-04-0-UI.s1 NCI_CGAP_FEO Human sapiens cDNA clone

```

```

ACCESSION      UI-H-FEO-bdn-c-04-0-UI 3', mRNA sequence.
VERSION      BO772836
KEYWORDS      GI:21981312
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 755)
JOURNAL      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@remail.nih.gov
Tissue Procurement: James Martin
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, Bento-Soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA-Yes.
FEATURES
location/Qualifiers
1..755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-FEO-bdn-c-04-0-UI"
/clone_lib="NCI_CGAP_FEO"
/tissue_type="Chondrosarcoma Cell line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_FEO is
a cDNA library containing the following tissue(s): a pool
of 3 chondrosarcoma cell lines (grade 2) The library was
constructed according to Bonaldi, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCTACGGAC. The cell lines was provided by Dr James Martin
of University of Iowa.
TAG_LIB=UI-H-FEO
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_SEQ=CGCTACGGAC"
BASE COUNT      207 a      141 c      115 g      290 t      2 others
ORIGIN
Query Match      20.1%; Score 676; DB 14; Length 765;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2930 ATGTATTTAGAAAATGATCATGATATGACCCAGAAATTAAGAAATTACATTG 2989
 |||||
 Db 455 ATGTATTTAGAAAATGATCATGATATGACCCAGAAATTAAGAAATTACATTG 396
 |||||
 QY 2990 TAAGAGAAACAAACAGCTGCAAAAAGAGTGGCCAGCTATCAATATAAAGCTGAAATCA 3049
 |||||
 Db 395 TAAGAGAAACAAACAGCTGCAAAAAGAGTGGCCAGCTATCAATATAAAGCTGAAATCA 336
 |||||
 QY 3050 TCTGCAATTTGCGCCAGCTGTTGGGACAAATGATGTTGTCACAAAGGCTTACATTGCTTT 3109
 |||||
 Db 335 TCTGCAATTTGCGCCAGCTGTTGGGACAAATGATGTTGTCACAAAGGCTTACATTGCTTT 276
 |||||
 QY 3110 GTCTCAAAATTAAGAAATTTTGTAGTGGTTTCAAAAATATTCACAAAGCAATACATA 3169
 |||||
 Db 275 GTCTCAAAATTAAGAAATTTTGTAGTGGTTTCAAAAATATTCACAAAGCAATACATA 216
 |||||
 QY 3170 AAAAGTGGTGAATTTACCTATACATTTCCCAATCTTGACATTCAGAAATGCTGTTAT 3229
 |||||
 Db 215 AAAAGTGGTGAATTTACCTATACATTTCCCAATCTTGACATTCAGAAATGCTGTTAT 156
 |||||
 QY 3230 TTAGTATGAGGATTAAGCACTTGTATGAAGATCTTTTAAATATCTATCATTTAAACAT 3289
 |||||
 Db 155 TTAGTATGAGGATTAAGCACTTGTATGAAGATCTTTTAAATATCTATCATTTAAACAT 96
 |||||
 QY 3290 TAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 3349
 |||||
 Db 95 TAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 36
 |||||
 QY 3350 TGATTTGTTTACTCTG 3365
 |||||
 Db 35 TGATTTGTTTACTCTG 20
 |||||

RESULT 3
 B0467983 1115 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6437921 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532884
 DEFINITION 5', mRNA sequence.
 ACCESSION B0467983
 VERSION B0467983.1 GI:18517025
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1115)
 NIH_MGC http://mgi.mcg.mcg.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rstraub@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM12216 row: n column: 21
 High quality sequence stop: 690.
 Location/Qualifiers
 1. 1115

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5532884"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomysarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NotI;
 Site-2: SalI; Cloned unidirectionally. Primer: oligo dT.
 Average insert size 2.1 kb."

BASE COUNT 326 a 221 c 218 g 277 t 73 others

ORIGIN

326 a 221 c 218 g 277 t 73 others

Query Match 19 68; Score 658; DB 13; Length 1115;
 Best Local Similarity 99.68; Pred. No. 0;
 Matches 808; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2555 TGAATTTAAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2614
 |||||
 Db 2 TGAATTTAAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 61
 |||||
 QY 2615 TGTGTCAGGCGCGTGTGTCAGGCGAGCTGATGAGACCTAGCTCTGTTATTTACA 2674
 |||||
 Db 62 TGTGTCAGGCGCGTGTGTCAGGCGAGCTGATGAGACCTAGCTCTGTTATTTACA 121
 |||||
 QY 2675 GTGGTTCAGGCGTATGCAAGATGAGACCTGATTAATGATTTCCAGAGAGAGATATA 2734
 |||||
 Db 122 GTGGTTCAGGCGTATGCAAGATGAGACCTGATTAATGATTTCCAGAGAGAGATATA 181
 |||||
 QY 2735 AAGCTATACATTTGTTCAAAATATGAAACAGAGAGATGATGATTAATGATTAAT 2794
 |||||
 Db 182 AAGCTATACATTTGTTCAAAATATGAAACAGAGAGATGATGATTAATGATTAAT 241
 |||||
 QY 2795 TACGATGCAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2854
 |||||
 Db 242 TACGATGCAAGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 301
 |||||
 QY 2855 ACAGAAATTAACCCATCACTATTAATCTTCTGCAAAAACAGAGAGAGAGAGAGAT 2914
 |||||
 Db 302 ACAGAAATTAACCCATCACTATTAATCTTCTGCAAAAACAGAGAGAGAGAGAGAT 361
 |||||
 QY 2915 CTGGGGAAGATATCATGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2974
 |||||
 Db 362 CTGGGGAAGATATCATGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 421
 |||||
 QY 2975 AGCAATTTACATTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3034
 |||||
 Db 422 AGCAATTTACATTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 481
 |||||
 QY 3035 TAAATGGAATATCATGCAAAATGTTGGCGAGCTTGGGGAACATATGATGATGATGAT 3094
 |||||
 Db 482 TAAATGGAATATCATGCAAAATGTTGGCGAGCTTGGGGAACATATGATGATGATGAT 541
 |||||
 QY 3095 GCTTAGATTTGCTGCTGCTCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAT 3154
 |||||
 Db 542 GCTTAGATTTGCTGCTGCTCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAT 601
 |||||
 QY 3155 CAAGAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3214
 |||||
 Db 602 CAAGAAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 661
 |||||
 QY 3215 CAGAAATGCTGTTATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3274
 |||||
 Db 662 CAGAAATGCTGTTATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 721
 |||||
 QY 3275 TATGAGTTAAACATTTATATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 3334
 |||||
 Db 722 TATGAGTTAAACATTTATATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 781
 |||||
 QY 3335 TAAGAAATTAATTAATGATTTTACTCTG 3365
 |||||
 Db 782 TAAGAAATTAATTAATGATTTTACTCTG 812
 |||||

RESULT 4

B0960157 870 bp mRNA linear EST 21-AUG-2002

LOCUS AGENCOURT_8949486 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6472264

DEFINITION 5', mRNA sequence.

ACCESSION B0960157

VERSION B0960157.1 GI:22375635

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 870)
 AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: L14M14005 row: 0 column: 17
 High quality sequence stop: 738.

FEATURES
 source
 1..870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NIH_MGC_71"
 /issue_type="telomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORE6; Site: 1; Nott; Site_2: Salt; Cloned unidirectionally. Primer: 01190 dt.
 Average insert size 2.1 kb."

BASE COUNT 319 a 131 c 204 g 214 t 2 others
 ORIGIN

Query Match 19.4% Score 652; DB 14; Length 870;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 702; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1976 ATTGAGAGAGTACATGAGGCGCTTACAAATTAAGACAAATTCGAAATTCATGATGCGT 2035
 DB 36 ATTGAGAGAGTACATGAGGCGCTTACAAATTAAGACAAATTCGAAATTCATGATGCGT 95
 QY 2036 ATACATCTTGAACCTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
 DB 96 ATACATCTTGAACCTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 155
 QY 2096 ATGATAGTATGAG 2155
 DB 156 ATGATAGTATGAG 215
 QY 2156 TAAAGAAACCTTTGAACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
 DB 216 TAAAGAAACCTTTGAACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 275
 QY 2216 ACAATTAATGTTGAAAGCGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2275
 DB 276 ACAATTAATGTTGAAAGCGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 335
 QY 2276 AATTAGAAATACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2335
 DB 336 AATTAGAAATACATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 395
 QY 2336 TCTTTACAAAAACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2395
 DB 396 TCTTTACAAAAACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 455
 QY 2396 AATTGCTGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2455
 DB 456 AATTGCTGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 515
 QY 2456 TCAAAACCATGACAG 2515
 DB 516 TCAAAACCATGACAG 575
 QY 2516 TCAATCTGCTATGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2575
 DB 576 TAAATCTGCTATGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 635

QY 2576 TTGTTATCCGTTATGCTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2635
 DB 636 TTGTTATCCGTTATGCTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 695
 QY 2636 CCAGAGCTGATGAG 2678
 DB 696 CCAGAGCTGATGAG 738

RESULT 5
 BM476961
 LOCUS BM476961
 DEFINITION AGENCOURT_6481569 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:555523
 5', mRNA sequence.
 ACCESSION BM476961
 VERSION BM476961.1 GI:18526003
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1046)
 AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: L14M12275 row: n column: 04
 High quality sequence stop: 622.

FEATURES
 source
 1..1046
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NIH_MGC_71"
 /issue_type="telomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORE6; Site: 1; Nott; Site_2: Salt; Cloned unidirectionally. Primer: 01190 dt.
 Average insert size 2.1 kb."

BASE COUNT 318 a 217 c 227 g 282 t 2 others
 ORIGIN

Query Match 18.4% Score 619; DB 13; Length 1046;
 Best Local Similarity 100.0%; Pred. No. 1.3e-306;
 Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 TCGAATGGTATTCACAG 221
 DB 1 TCGAATGGTATTCACAG 60
 QY 232 GTGAAGTATGATCCAG 291
 DB 61 GTGAAGTATGATCCAG 120
 QY 292 GTGAAGGAGAGATGAG 351
 DB 121 GTGAAGGAGAGATGAG 180
 QY 352 CTGCTGAGACCTTGGAG 411
 DB 181 CTGCTGAGACCTTGGAG 240
 QY 412 GCCCTCGGAG 471

Db 241 GGCCTCGGAGAACCGGAGCCCTCTGGCCGCCGCTACATGAACTGATGCTACAGGAG 300

QY 472 TTGCCCTCCATCGTTGAGAACGCTATGATGATATCTGAACTGCTGTAACCTCCTT 531

Db 301 TTGCCCTCCATCGTTGAGAACGCTATGATGATATCTGAACTGCTGTAACCTCCTT 360

QY 532 CAGCCCACTCTGTTGGACAGCTTGTAGTAGAGAGCTTGGATAGAGCTGAGAG 591

Db 361 CAGCCCACTCTGTTGGACAGCTTGTAGTAGAGAGCTTGGATAGAGCTGAGAG 420

QY 592 GAAGCTGTGACAAATTGAGACAGAAACCGGATTGCTGTCGAGAAACAAATGGAATGAA 651

Db 421 GAAGCTGTGACAAATTGAGACAGAAACCGGATTGCTGTCGAGAAACAAATGGAATGAA 480

QY 652 TAGAGGTAGAGAGCTTCTAAAGAGATGTCGAGAAAGAAAGAAAGAAAGAAAGTTCATTT 711

Db 481 TAGAGGTAGAGAGCTTCTAAAGAGATGTCGAGAAAGAAAGAAAGAAAGTTCATTT 540

QY 712 CTGAATGTTCTGTCGACAAAGAGAAACAAATGATGATGATGATGATGATGATGATGAT 771

Db 541 CTGAATGTTCTGTCGACAAAGAGAAACAAATGATGATGATGATGATGATGATGATGAT 600

QY 772 TGCTCAGAAAGCAATGACG 790

Db 601 TGCTCAGAAAGCAATGACG 619

RESULT 6
BF337464 1035 bp mRNA lineal EST 22-NOV-2000
LOCUS BF337464 60203519591 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183126
DEFINITION 5', mRNA sequence.
ACCESSION BF337464
VERSION BF337464.1 GI:11283715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1035)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9498 row: n column: 23
High quality sequence start: 5
High quality sequence stop: 695.
Location/Qualifiers
1..1035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4183126"
/clone_lib="NCI_CGAP_Brn64"
/tissue_type="globlastoma with PCR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-Sport6; Site:1: Not;
Site:2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 353 a 221 c 243 g 218 t
ORIGIN

Query Match 18.0%; Score 607; DB 2: Length 1035;
Best Local Similarity 99.8%; Pred. No. 2e-30;
Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1049 AGAATGTGGCAGCAAGACATCCCGGAGCCAGCAAGCTCCAGCTGAGGCTTAAATATG 1108

Db 4 AGAATGTGGCAGCAAGACATCCCGGAGCCAGCAAGCTCCAGCTGAGGCTTAAATATG 63

QY 1109 AAGTTGCCACCGACCTTGGAGGAGAAATATCATCATCTGCTGCTGAATGAGATG 1168

Db 64 AAGTTGCCACCGACCTTGGAGGAGAAATATCATCATCTGCTGCTGCTGAATGAGATG 123

QY 1169 GAAATACAGAGTGGCTTTACATTCGCAAGATCATTGACAGAGAGAGAAATAT 1228

Db 124 GAAATACAGAGTGGCTTTACATTCGCAAGATCATTGACAGAGAGAGAAATAT 183

QY 1229 CTGAGCCGTGAAAGTTATAGTCTTCGATTAAGTACTGCTGATGAAATATCTTCC 1288

Db 184 CTGAGCCGTGAAAGTTATAGTCTTCGATTAAGTACTGCTGATGAAATATCTTCC 243

QY 1289 GCAAGAGCTTCCACCATTTTGAAGAAATGCTATGCTGATGATTAATATATGATA 1348

Db 244 GCAAGAGCTTCCACCATTTTGAAGAAATGCTATGCTGATGATTAATATATGATA 303

QY 1349 CCAAGTGAATATATCATTTCCAGAGTTGCAAGTCTGATATTTATATAGTAG 1408

Db 304 CCAAGTGAATATATCATTTCCAGAGTTGCAAGTCTGATATTTATATAGTAG 363

QY 1409 CTCAATCTCTGAAAGACCTCCCTTAAACCTGGAATGGAAGATGCTGCTTCAAT 1468

Db 364 CTCAATCTCTGAAAGACCTCCCTTAAACCTGGAATGGAAGATGCTGCTTCAAT 423

QY 1469 TGTGAGCTTTCCCTCATTTATCATTTGATGATGATGATGATGATGATGATGATGAT 1528

Db 424 TGTGAGCTTTCCCTCATTTATCATTTGATGATGATGATGATGATGATGATGATGAT 483

QY 1529 ATATATCATATGAGGATTTATTTGATGCAAGAGTTGAAATGAAATATGACT 1588

Db 484 ATATATCATATGAGGATTTATTTGATGCAAGAGTTGAAATGAAATATGACT 543

QY 1589 AAAACAAACAGTATGCTCCCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1648

Db 544 AAAACAAACAGTATGCTCCCTGCTGATGATGATGATGATGATGATGATGATGATGAT 603

QY 1649 GAGGGGCGCAGAGAGCAAGCAAGCTGAAGACATTTAAACTATGTCATAT 1706

Db 604 GAGGGGCGCAGAGAGCAAGCAAGCTGAAGACATTTAAACTATGTCATAT 661

RESULT 7
BF983236 729 bp mRNA lineal EST 23-JAN-2001
LOCUS BF983236 602305873P1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:417083 5'
DEFINITION mRNA sequence.
ACCESSION BF983236
VERSION BF983236.1 GI:12386048
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 729)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10097 row: a column: 20
High quality sequence stop: 665.

FEATURES
source
Location/Qualifiers
1. 729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4397083"
/issue_type="NIH_MGC_88"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPOK79;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC binary."

BASE COUNT 240 a 137 c 184 g 168 t
ORIGIN

Query Match 16.0%; Score 539; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.9e-265;
Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 586 GAGGAGAACTGTTGACATTTGAAGACAGAAACCGATTGCTGTCGCA/AAACATGGA 645
Db 1 GAGGAGAACTGTTGACATTTGAAGACAGAAACCGATTGCTGTCGCA/AAACATGGA 60
QY 646 AATGATAGTGTGAAGACACTACTATAAAGATTTGCGAAGAAGAAGTGTCTCT 705
Db 61 AATGATAGTGTGAAGACACTACTATAAAGATTTGCGAAGAAGAAGTGTCTCT 120
QY 706 GCATTTCGATGTTCTTCGTCGAAGAGAAACATGAACTGTCGAAGATTACAGGC 765
Db 121 GCATTTCGATGTTCTTCGTCGAAGAGAAACATGAACTGTCGAAGATTACAGGC 180
QY 766 TCTGATTCCTCAGAAAGCATGACAGATTTGAGATTTATCA/AAAGTTGATGCTCTCAA 825
Db 181 TCTGATTCCTCAGAAAGCATGACAGATTTGAGATTTATCA/AAAGTTGATGCTCTCAA 240
QY 826 GTGGAAGAGCAACTCTTTCACCAAGATTCAGCAAAATCTGGAAGAGAGGTCTGGGGC 885
Db 241 GTGGAAGAGCAACTCTTTCACCAAGATTCAGCAAAATCTGGAAGAGAGGTCTGGGGC 300
QY 886 ATGAGATTAACATCATGATCAATCTTTTGCAGATTTCTTCTAGTTT/AGATTCAGAC 945
Db 301 ATGAGATTAACATCATGATCAATCTTTTGCAGATTTCTTCTAGTTT/AGATTCAGAC 360
QY 946 ACAAGTTTGGCAGAGCAAGAGTGTAGCTGTTAGATGAAGT/TTGACAT/AAACAGCAAC 1005
Db 361 ACAAGTTTGGCAGAGCAAGAGTGTAGCTGTTAGATGAAGT/TTGACAT/AAACAGCAAC 420
QY 1006 ATGGGCAATGATTCAGGACCATGGAAGTATTCAGATGA/AGATTC/GGCAGCAGA 1065
Db 421 ATGGGCAATGATTCAGGACCATGGAAGTATTCAGATGA/AGATTC/GGCAGCAGA 480
QY 1066 GCATCCCGGAGCGAGCAAGCTCCAGCTGAGCAAAAT/TAAGTTTTCAGCCAGC 1124
Db 481 GCATCCCGGAGCGAGCAAGCTCCAGCTGAGCAAAAT/TAAGTTTTCAGCCAGC 539

RESULT 8
BM467774 1239 bp mRNA linear EST 05-FEB-2002
LOCUS ACBNCOURL6437783 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532660
DEFINITION 5', mRNA sequence.
ACCESSION BM467774
VERSION BM467774.1 GI:18516816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1239)
NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the J.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM12216 row: e column: 13
High quality sequence stop: 505.

FEATURES
source
Location/Qualifiers
1. 1239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5532660"
/clone_lib="NIH_MGC_71"
/issue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPOK79; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primers: Oligo dT.
Average insert size 2.1 kb."

BASE COUNT 393 a 285 c 243 g 318 t
ORIGIN

Query Match 15.3%; Score 514; DB 13; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.5e-252;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 736 AACATGAACCTTCCCAAGGTTACAGAGCTGATTCGTCGAAGCAAT/TAAGATT 795
Db 4 AACATGAACCTTCCCAAGGTTACAGAGCTGATTCGTCGAAGCAAT/TAAGATT 63
QY 796 GAGATTTATCAGAGTGTAGTGTCTCAAGTGAAGAGCAACTTTTCA/CAACATT 855
Db 64 GAGATTTATCAGAGTGTAGTGTCTCAAGTGAAGAGCAACTTTTCA/CAACATT 123
QY 856 CAGCCAAATGTGAGAGAGAGGTCTGGGCGATGAGAAATATCATCATCAAT/ACCTTT 915
Db 124 CAGCCAAATGTGAGAGAGAGGTCTGGGCGATGAGAAATATCATCATCAAT/ACCTTT 183
QY 916 GCAGATTCCTCTAGTTCAGATCAGACACAAAGTTTGACAGAGAGT/AGAGT/AGC 975
Db 184 GCAGATTCCTCTAGTTCAGATCAGACACAAAGTTTGACAGAGAGT/AGAGT/AGC 243
QY 976 TTAGATGAAGAGTCTTGACATAACAGCAATGGGAGTATTCAGGACCA/AGAAAT 1035
Db 244 TTAGATGAAGAGTCTTGACATAACAGCAATGGGAGTATTCAGGACCA/AGAAAT 303
QY 1036 GATTGAGATTAAGAGATGTGGCAGAGAGCATCCCGGAGCGAGAACT/TAAGT/AGG 1095
Db 304 GATTGAGATTAAGAGATGTGGCAGAGAGCATCCCGGAGCGAGAACT/TAAGT/AGG 363
QY 1096 CTTTACCAATGTGAGAGTGGCCAGCCAGCTTGGAAGGGAAGAT/TAAGT/AGG 1155
Db 364 CTTTACCAATGTGAGAGTGGCCAGCCAGCTTGGAAGGGAAGAT/TAAGT/AGG 423
QY 1156 CCTACAGGAGTGAAGAAACCAAGAGTGGCTTTTACATTTGCCAAGGATCA/TAAGT/AGG 1215
Db 424 CCTACAGGAGTGAAGAAACCAAGAGTGGCTTTTACATTTGCCAAGGATCA/TAAGT/AGG 483
QY 1216 AACAAAAAGCATGTGAGCTGTGAAAAAGTTATAG 1249
Db 484 AACAAAAAGCATGTGAGCTGTGAAAAAGTTATAG 517

RESULT 9
BE882040 1041 bp mRNA linear EST 20-OCT-2000
LOCUS BE882040
DEFINITION 601505326F2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532660
ACCESSION BE882040
VERSION BE882040.1 GI:10330816

| KEYWORDS | EST. |
|-----------------------|---|
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | 1 (bases 1 to 1041) |
| TITLE | NIH-MGC http://mgc.ncl.nih.gov/ . |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapds-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM9716 row: d column: 07 High quality sequence stop: 586. Location/Qualifiers |
| FEATURES | 1..1041 |
| source | /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="3906774" /clone_lib="NIH-MGC_71" /tissue_type="leiomyosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pCMV-SpOx5; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. |
| BASE COUNT | 385 a 258 c 223 g 175 t |
| ORIGIN | |
| Query Match | 14.9%; Score 501; DB 12; Length 1041; |
| Best Local Similarity | 100.0%; Pred. No. 6,9e-246; |
| Matches 501; | Conservative 0; Mismatches 0; Indels 0; Gaps 0 |
| OY 2010 | TGACCAATTGCAATGATGATGCGTATACGTATTTGAACCTTCCTATATATGAACAGAA 2069 |
| Db 89 | TGACCAATTGCAATGATGATGCGTATACGTATTTGAACCTTCCTATATATGAACAGAA 148 |
| OY 2070 | AGATAAGAAGTTGGAGCATGGAAGATATAGTATGAGGCTGGATATGAGATATG 2129 |
| Db 149 | AGATAAGAAGTTGGAGCATGGAAGATATAGTATGAGGCTGGATATGAGATATG 208 |
| OY 2130 | TGATGATGATGAAGATGAGATGATTTAAAGAACTTTGAACCTGGATATGAACAGATAG 2189 |
| Db 209 | TGATGATGATGAAGATGAGATGATTTAAAGAACTTTGAACCTGGATATGAACAGATAG 268 |
| OY 2190 | ATTTCATGATCTTATTTTGTGAAAACAAATAAATGTTGAAAGAGCTGTGAAAACCC 2249 |
| Db 269 | ATTTCATGATCTTATTTTGTGAAAACAAATAAATGTTGAAAGAGCTGTGAAAACCC 328 |
| OY 2250 | AGAAATATGAAAATGAAAAGCTGACCACAAATAGAAATATACATATAGAGTATATACAG 2309 |
| Db 329 | AGAAATATGAAAATGAAAAGCTGACCACAAATAGAAATATACATATAGAGTATATACAG 388 |
| OY 2310 | GACTAGAGCAATAGACACGAGGAATATCTTTACAAAAACACACAGAGTATATGCGCT 2369 |
| Db 389 | GACTAGAGCAATAGACACGAGGAATATCTTTACAAAAACACACAGAGTATATGCGCT 448 |
| OY 2370 | TTCCCGAGTGATTTACTGAAAATGAAAAATTTGCTGAGTAGAGAGTGAAGCTGCACATCT 2429 |
| Db 449 | TTCCCGAGTGATTTACTGAAAATGAAAAATTTGCTGAGTAGAGAGTGAAGCTGCACATCT 508 |
| OY 2430 | GATTGAGCTGACACAGACAGTGAATCAAAACCATGACACATATGAATGAATGAAGAAGT 2489 |
| Db 509 | GATTGAGCTGACACAGACAGTGAATCAAAACCATGACACATATGAATGAATGAAGAAGT 568 |
| OY 2490 | CATTAGTAATATTTGCACATGG 2510 |
| Db 569 | CATTAGTAATATTTGCACATGG 589 |

| | |
|-----------------------|---|
| RESULT 10 | |
| LOCUS | B0016388/c |
| DEFINITION | B0016388 |
| ACCESSION | B0016388 |
| VERSION | B0016388.1 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| TITLE | NCI-CGAP clone http://www.ncbi.nlm.nih.gov/ncicgap. |
| JOURNAL | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), |
| COMMENT | Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgabs@remail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Seq primer: M13 FORWARD POLYA-Yes. |
| FEATURES | Location/Qualifiers |
| SOURCE | 1..619 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image:5887882" /clone_lib="NCI-CGAP_DTI" /tissue_type="Metastatic Chondrosarcoma" /stage="Adult" /lab_host="DH10B (Life Technologies)" /note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) With a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI-CGAP DTIs is a normalized cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung." The library was constructed according to Lynch39, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTTCGG. TAG_LIB=UI-H-DTI TAG_TISSUE=Lung metastatic chondrosarcoma TAG_SEQ=AACTGTTCGG" |
| BASE COUNT | 172 a 109 c 91 g 247 t |
| ORIGIN | |
| Query Match | 14.8%; Score 498; DB 14; Length 619 |
| Best Local Similarity | 99.7%; Pred. No. 2.3e-244; |
| Matches 598: | Conservative 0; Mismatches 2; Indels 0; Gaps 0 |
| QY | 2766 AGAGAGATGTGCCTAACAATTGGGAATGCAAGAATATCATTAATGAAAT 2825 |
| DB | 619 AGAGAGATGTGCTCATAGATTTTGGAATTACAGATGCAAAATATATATGAAAT 560 |
| QY | 2826 GAAGAACAGAGAAATTTGCCAAGCATTACAAGATAACCATCACTAATTAATTTCT 2885 |
| DB | 559 GAAGCCAAAGAAATATATGCCAAGCATTAACAAGATAACCATCATCAATAATTTCT 500 |
| QY | 2886 TTGCAGAAAACTGCATGTCTAGCCTGTTCTGGGGAAGATATTCATGTAAATTAATAAT 2945 |
| DB | 499 TTGCAGAAAACCTGCTGCTAGCCTGTTCTGGGGAAGATATTCATGTAAATTAATAAT 440 |

QY 2946 GCATCAGCTCATATGACCCCGAGATTCAGAACTTACATGTAAAGTAAACAAAGC 3005
 DB 439 GCATCAGCTCATATGACCCCGAGATTCAGAACTTACATGTAAAGTAAACAAAGC 380
 QY 3006 ACTGCAAAAGAGTGTCCGACATCAATTAATGTGCAATATCTGCAATGTGCCA 3065
 DB 379 ACTGCAAAAGAGTGTCCGACATCAATTAATGTGCAATATCTGCAATGTGCCA 320
 QY 3066 GCGTGGGAGCAATGATGAGCGCAAAAGGCTTAGATTGCTGTGTCGCAAAATAGGAA 3125
 DB 319 GCGTGGGAGCAATGATGAGCGCAAAAGGCTTAGATTGCTGTGTCGCAAAATAGGAA 260
 QY 3126 TTTGTAGTGTGTTTCAAAATTAATCAACAAGAACAAATTAATAAGCTGGTAGAAT 3185
 DB 259 TTTGTAGTGTGTTTCAAAATTAATCAACAAGAACAAATTAATAAGCTGGTAGAAT 200
 QY 3186 ACCTATGCAATTTCCCATCTTGTGATTCAGATTCGCTTATTTAGTCAATGAGATTA 3245
 DB 199 ACCTATGCAATTTCCCATCTTGTGATTCAGATTCGCTTATTTAGTCAATGAGATTA 140
 QY 3246 GCATTTGATGATGATCTTTTAAATACATCAATTAATTAATGATTTATGATTT 3305
 DB 139 GCATTTGATGATGATCTTTTAAATACATCAATTAATTAATGATTTATGATTT 80
 QY 3306 AATGATTTATGATGATCTTCAAGACATGACATTAATTAATGATTTTATGATTT 3365
 DB 79 AATGATTTATGATGATCTTCAAGACATGACATTAATTAATGATTTTATGATTT 20

RESULT 11
 LOCUS BF686405 781 bp mRNA linear EST 22-DEC-2000
 DEFINITION 602143786F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4304805 5',
 mRNA sequence.
 ACCESSION BF686405
 VERSION BF686405.1 GI:11971813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 781)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Smith Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LCM1170 row: 1 column: 22
 High quality sequence stop: 731.
 Location/Qualifiers
 1..781

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4304805"
 /clone_id="NIH_MGC_46"
 /tissue_type="leiomyosarcoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pGB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(d). Size selected >0.04µ for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC

BASE COUNT 235 a 168 c 204 g 174 t
 ORIGIN

Query Match 14.5%; Score 489; DB 12; Length 761;
 Best Local Similarity 99.8%; Pred. No. 1e-239;
 Matches 609; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 392 GGAATGCGAATTCGTTGAGAGGCGCTCGGAGAACCGGACCCCTGTGGCGTTCCTATA 451
 DB 12 GGAATGCGAATTCGTTGAGAGGCGCTCGGAGAACCGGACCCCTGTGGCGTTCCTATA 71
 QY 452 TGAACCTGAGCTCAGGAGCTGCTCCCTGCTCAGCTTTGAGAAAGCTCATGATTAAT 511
 DB 72 TGAACCTGAGCTCAGGAGCTGCTCCCTGCTCAGCTTTGAGAAAGCTCATGATTAAT 131
 QY 512 TCCAGCTGCTGAACCTGCTCAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571
 DB 132 TCCAGCTGCTGAACCTGCTCAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 190
 QY 572 TGGATTAAGTCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 631
 DB 191 TGGATTAAGTCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 250
 QY 632 CAGAAACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 691
 DB 251 CAGAAACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 310
 QY 692 AAAAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
 DB 311 AAAAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370
 QY 752 AAGAGTTACAGGCTGCTGATTCGCTCAGAAAGCAATGAGATTTGAGATTTATATTAAG 811
 DB 371 AAGAGTTACAGGCTGCTGATTCGCTCAGAAAGCAATGAGATTTGAGATTTATATTAAG 430
 QY 812 TTGATGCTCTCAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 871
 DB 431 TTGATGCTCTCAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 490
 QY 872 AGGAGGCTGGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 931
 DB 491 AGGAGGCTGGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 550
 QY 932 TTTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991
 DB 551 TTTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 610
 QY 992 GACATTAACG 1001
 DB 611 GACATTAACG 620

RESULT 12
 LOCUS A1718277 536 bp mRNA linear EST 10-JUN-1999
 DEFINITION as51e06.x1 Barslead aorta HPLK6 Homo sapiens cDNA clone
 IMAGE:2320738 3', mRNA sequence.
 ACCESSION A1718277
 VERSION A1718277.1 GI:5035533
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 536)

REFERENCE
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gels, J., Jost, S.,
 Krizman, D., Kudaba, T., Lacy, M., Le, N., Lennon, G., Mitra, M., Martin,
 J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Wiesting, B.,
 White, Y., Wylie, T., Waterston, R., and Wilson, R.
 Title WASHU-NCI human EST Project
 Journal Unpublished (1997)
 Comment Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 481.

FEATURES

source
1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2320738"
/clone_id="Barstead aorta HPLRB6"
/sex="male"
/dev_stage="adult, age 64"
/lab_host="DH10B (phage resistant)"
/note="Organ: aorta; Vector: p173D-Pac (Pharmacia) with a
modified polylinker; Site:1: EcoRI; Site:2: NotI; 1st
strand cDNA was primed with a Not I - c190(d1) primer (5'
TGTACCAATCTGAGTGGAGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
(5' AATTCGATCGAAC 3' and 5' GTTGGATCGG 3'), digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."

BASE COUNT 157 a 93 c 84 g 202 t
ORIGIN

Query Match 14.3%; Score 482; DB 9; Length 736;
Best Local Similarity 99.8%; Pred. No. 3.8e-246;
Matches 532; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2833 AAGAAATATTTGGCAACATTACAGAAATACCATCATATATCTCTTTTGGCAA 2892
|||||
Db 536 AAGAAATATTTGGCAACATTACAGAAATACCATCATATATCTCTTTTGGCAA 477
2893 AACTGAGTGTCTAGCTGTCTGGGAGATATCCATGTAATGAGAAATGATCAG 2952
Db 476 AACTGAGTGTCTAGCTGTCTGGGAGATATCCATGTAATGAGAAATGATCAG 417
2953 GTCAATATGACCCGAGAAATTCAGAACTTTACATTTGAAGAAACAAACACTGCA 3012
|||||
Db 416 GTCAATATGACCCGAGAAATTCAGAACTTTACATTTGAAGAAACAAACACTGCA 357
3013 AAGAAGTGTGGCACTATCAATTAATGTGAATCATCTGTAATGTGGCTAGCTTGG 3072
|||||
Db 356 AAGAAGTGTGGCACTATCAATTAATGTGAATCATCTGTAATGTGGCTAGCTTGG 297
3073 GGAACAATGATGGTGCACAAAGGCTTAGATTGCTTGTCTAAATATACCAATTTGTA 3132
|||||
Db 296 GGAACAATGATGGTGCACAAAGGCTTAGATTGCTTGTCTAAATATACCAATTTGTA 237
3133 GTGTTTTCAAAAATTAATTCACAAAGAAACAATACAAAATTTGGTAGAATTACTATG 3192
|||||
Db 236 GTGTTTTCAAAAATTAATTCACAAAGAAACAATACAAAATTTGGTAGAATTACTATG 177
3193 ACATTTCGAATCTTGACTATTCAGAAATGCTGTATTATTAGTAGAGAGTAGACTGTG 3252
Db 176 ACATTTCGAATCTTGACTATTCAGAAATGCTGTATTATTAGTAGAGAGTAGACTGTG 117
3253 ATTGAAGATCTTTTAAATATCACTAGTAAACATTTAATATGATTTATATGAT 3312
Db 116 ATTGAAGATCTTTTAAATATCACTAGTAAACATTTAATATGATTTATATGAT 57
3313 TCATTATGCTACAGAACTGACATAGAAATCAATAAATGATTTTATCTG 3365
Db 56 TCATTATGCTACAGAACTGACATAGAAATCAATAAATGATTTTATCTG 4

RESULT 13
AM589567/c

LOCUS AM589567 475 bp mRNA linear EST: 22-MAR-2000
DEFINITION x012b12.x1 NCI CGAP Ut3 Homo sapiens cDNA clone IMAGE:2320738 3'

ACCESSION AM589567
VERSION AM589567.1 GI:7276681

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo;
1 (bases 1 to 475)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapus-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
image.lnl.gov/image/html/lresources.shtml

Possible reversed clone: polyT not found
Seq primer: -400P from Glibco
High quality sequence stop: 407.

FEATURES

source
1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2703743"
/clone_id="NCI-CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT1; Site:1: SalI;
Site:2: NotI. Cloned unidirectionally. Primer: c190 d1.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

BASE COUNT 142 a 88 c 71 g 174 t
ORIGIN

Query Match 14.1%; Score 473; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.6e-231;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2893 AACTGAGTGTGCTAGCTGTCTGGGAGATATCCATGTAATTGAGAAATATATAC 2952
|||||
Db 475 AACTGAGTGTGCTAGCTGTCTGGGAGATATCCATGTAATTGAGAAATATATAC 416
2953 GTCAATATGACCCGAGAAATTCAGAACTTTACATTTGAAGAAACAAATTTGTA 3012
|||||
Db 415 GTCAATATGACCCGAGAAATTCAGAACTTTACATTTGAAGAAACAAATTTGTA 356
3013 AAGAAGTGTGGCACTATCAATTAATGTGAATCATCTGCAATATGCTTGG 3072
|||||
Db 355 AAGAAGTGTGGCACTATCAATTAATGTGAATCATCTGCAATATGCTTGG 296
3073 GGAACAATGATGGTGCACAAAGGCTTAGATTGCTTGTCTAAATATACCAATTTGTA 3132
|||||
Db 295 GGAACAATGATGGTGCACAAAGGCTTAGATTGCTTGTCTAAATATACCAATTTGTA 236
3133 GTGTTTTCAAAAATTAATTCACAAAGAAACAATACAAAATTTGGTAGAATTACTATG 3192
|||||
Db 235 GTGTTTTCAAAAATTAATTCACAAAGAAACAATACAAAATTTGGTAGAATTACTATG 176
3193 ACATTTCGAATCTTGACTATTCAGAAATGCTGTATTATTAGTAGAGAGTAGACTGTG 3252
Db 175 ACATTTCGAATCTTGACTATTCAGAAATGCTGTATTATTAGTAGAGAGTAGACTGTG 116

| QY | 3253 | ATTGAAGCTTCTTTTAAATCTATCACTTTAAACATTAATGAATTAATGAT | 3312 |
|-----------------------|--|---|------------------------------------|
| Db | 115 | ATTGAAGCTTCTTTTAAATCTATCACTTTAAACATTAATGAATTAATGAT | 56 |
| QY | 3313 | TCATTATCTACAGAACTGACATTAAGATCAATAATGATGTTTACTG | 3365 |
| Db | 55 | TCATTATCTACAGAACTGACATTAAGATCAATAATGATGTTTACTG | 3 |
| RESULT 14 | | | |
| AI922705/c | | | |
| LOCUS | AI922705 | 521 bp | mRNA |
| DEFINITION | w011f08.x1 NCI-CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2455047 3', mRNA sequence. | | |
| ACCESSION | AI922705 | | |
| VERSION | AI922705.1 | GI:5658669 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | |
| AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. | | |
| TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index | | |
| JOURNAL | Unpublished (1997) | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Life Technologies catalog #: 11548-013 DNA sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: www.bio.llnl.gov/dbp/image/image.html Seq primer: -400p from Glcbo High quality sequence stop: 410. Location/Qualifiers | | |
| FEATURES | source | | |
| | 1..521 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone_image="2455047" | | |
| | /clone_id="NCI-CGAP_Pan1" | | |
| | /tissue_type="adenocarcinoma" | | |
| | /lab_host="DH10B" | | |
| | /note="Organ: pancreas; Vector: pCMV-SV40; Site: 1; Salt: Site-2; Note: Cloned unidirectionally, primer: oligo dt. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013" | | |
| BASE COUNT | 154 a | 90 c | 82 g |
| ORIGIN | 195 t | | |
| Query Match | 13.8% | Score 466; | DB % Length 721; |
| Best Local Similarity | 99.8% | Pred. No. 6.4e-22H; | |
| Matches | 516; | Conservative | 0; Mismatches 1; Indels 0; Gaps 0; |
| QY | 2849 | AGCATTACAGAAATACCCATCACTAATTAATCTTTCCTTTCGAAAACAGTCACTGCTAG | 2908 |
| Db | 521 | AGCATTACAGAAATACCCATCACTAATTAATCTTTCCTTTCGAAAACAGTCACTGCTAG | 462 |
| QY | 2909 | CCTGTTGGGGAAATATATTCATATGTAATTTGAGAAATGATCATGTCGAAATATGACCCAG | 2968 |
| Db | 461 | CCTGTTGGGGAAATATATTCATATGTAATTTGAGAAATGATCATGTCGAAATATGACCCAG | 402 |
| QY | 2969 | AATTCAGAGAACTTACATTTGTATAGAGAAACAAAGCAG | 3028 |
| Db | 401 | AATTCAGAGAACTTACATTTGTATAGAGAAACAAAGCAGTGTAAAGAACTGTGCGACT | 342 |
| QY | 3029 | ATCAAAATAATGGTAAATCATCTGCAAAATGTGGCCAGGCTTGGGAAATATGATGTGC | 3088 |
| Db | 341 | ATCAAAATAATGGTAAATCATCTGCAAAATGTGGCCAGGCTTGGGAAATATGATGTGC | 282 |
| QY | 3089 | ACAAAGGCTTACATTTGCTGCTGTCGAAATTAAGAAATTTTCACTGTTTCGAAATA | 3148 |
| Db | 281 | ACAAAGGCTTACATTTGCTGCTGTCGAAATTAAGAAATTTTCACTGTTTCAAAATA | 222 |

| | | | |
|-----------------------|---|---|--------------------|
| QY | 3149 | ATTGACGAAACAAACAAATACGACAAAAGGCGGAGAGATTCACCATACATTCCTGCAATCTTG | 3208 |
| Db | 221 | ATTGACGAAACAAACAAATACGACAAAAGGCGGAGAGATTCACCATTCATTCACATTTTCAATCTTG | 162 |
| QY | 3209 | ACTATTCGAAATGCGTGTATTATTAGTGTAGTGAGGATTTAGCAGCTTGATTGAAATGTTTGA | 3268 |
| Db | 161 | ACTATTCGAAATGCGTGTATTATTAGTGTAGTGAGGATTTAGCAGCTTGATTGAAATGTTTGA | 102 |
| QY | 3269 | AAATACATCTACGATTTAAACATTTTATATGATATATGATTTAATGTATTTCATTATCTACGATAA | 3328 |
| Db | 101 | AAATACATCTACGATTTAAACATTTTATATGATATATGATTTAATGTATTTCATTATCTACGATAA | 42 |
| QY | 3329 | CTGACATAGAGATCATATAAATGATTTGTTTACTCTG | 3365 |
| Db | 41 | CTGACATAGAGATCATATAAATGATTTGTTTACTCTG | 5 |
| RESULT | 15 | | |
| LOCUS | BE940626 | 546 bp | mRNA |
| DEFINITION | BC3-UTR00064-170800-021-e06 | UTR0064 | Homo sapiens cDNA |
| VERSION | BE940626 | | |
| KEYWORDS | EST | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 546) Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, D.F., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, G.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J. | | |
| TITLE | Shotgun sequencing of the human transcriptome with 3Kf-expressed sequence tags | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. | 97 (7) | 3491-3496 (2000) |
| MEDLINE | 20202663 | | |
| COMMENT | Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL: (http://www.ludwig.org.br/scripts/gethtml2.pl?file=BC3-UTR0064-170800-021-e06&3=2000-08-17&4=1) Seq primer: puc 18 forward High quality sequence start: 53 High quality sequence stop: 542. Location/Qualifiers 1..546 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="UTR0064" /dev_stage="Adult" /note="Organ: uterus; tumor: Vector: puc18; Site: 1; Smal: Site: 2; Smal: A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." | | |
| BASE COUNT | 205 a | 86 c | 102 g |
| ORIGIN | 153 t | | |
| Query Match | 13.8% | Score 463; | DB 12; Length 446; |
| Best Local Similarity | 100.0%; | Pred No. 2.3e-226; | |

Matches 463: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 2844 TGCCAGAGATTACAGATTAACCCATCTATTAATCTTCTTGGCAAAATCTGCAGTGT 2903
      |||||||
Db 77 TGCCAGAGATTACAGAAATTAACCCATCTATTAATCTTCTTGGCAAAATCTGCAGTGT 136
      |||||||
OY 2904 GCTAGCCGTCTGGGGAAGATATCCATGTATGAGAAATGCATCAGCTCATATATGAC 2963
      |||||||
Db 137 GCTAGCCGTCTGGGGAAGATATCCATGTATGAGAAATGCATCAGCTCATATATGAC 196
      |||||||
OY 2964 CCCAGAAATTCAGGAACCTTACATTTGTAAGAAATCAAGCACTGCAGAAATGAAGTGTGC 3023
      |||||||
Db 197 CCCAGAAATTCAGGAACCTTACATTTGTAAGAAATCAAGCACTGCAGAAATGAAGTGTGC 256
      |||||||
OY 3024 CGACTATCAATTAATGTGTAATCATCTGCAAAATGTGCGAGGCTTGGGACAAATGAT 3083
      |||||||
Db 257 CGACTATCAATTAATGTGTAATCATCTGCAAAATGTGCGAGGCTTGGGACAAATGAT 316
      |||||||
OY 3084 GGTGCACAAAGGCTAGATTGGCTTGTCTCAAAATAGGAATTTTGTAGTGTTCAA 3143
      |||||||
Db 317 GGTGCACAAAGGCTAGATTGGCTTGTCTCAAAATAGGAATTTTGTAGTGTTCAA 376
      |||||||
OY 3144 AAATTAATTCACAAAGAAACATATACAAAAGTGGGTAGAAATACCTATCAATTTCCCAA 3203
      |||||||
Db 377 AAATTAATTCACAAAGAAACATATACAAAAGTGGGTAGAAATACCTATCAATTTCCCAA 436
      |||||||
OY 3204 TCTTGACTATTCAGATGCTGTTTATTAGTAGAGATAGCACTTGATTTGAAGATTC 3263
      |||||||
Db 437 TCTTGACTATTCAGATGCTGTTTATTAGTAGAGATAGCACTTGATTTGAAGATTC 496
      |||||||
OY 3264 TTTTAAATACATATCAGTTAAACATTTAATAGATTATGATTA 3306
      |||||||
Db 497 TTTTAAATACATATCAGTTAAACATTTAATAGATTATGATTA 539
      |||||||
```

Search completed: May 17, 2003, 11:01:43
Job time : 4534 secs

...

GenCore version 5.1.4.B5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 13:25:51 ; Search time 115 Seconds
(without alignments)
8973.629 Million (all) updates/sec

Title: US-09-515-363C-1

Perfect score: 3365
Sequence: 1 gcgcgcgcgcctgagagcc.....aaatgattgtttactctg 3365

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882.1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents.NA.*

1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PCUTUS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--|
| 1 | 129 | 3.8 | 301 | 1 | US-08-143-576-7 Sequence 7, Appl |
| 2 | 108 | 3.2 | 193 | 1 | US-08-143-576-6 Sequence 6, Appl |
| 3 | 22 | 0.7 | 169998 | 4 | US-09-676-610B-24 Sequence 24, Appl |
| 4 | 20 | 0.6 | 139 | 4 | US-09-345-882-12 Sequence 12, Appl |
| 5 | 20 | 0.6 | 1418 | 4 | US-09-111-470-7 Sequence 7, Appl |
| 6 | 20 | 0.6 | 6002 | 4 | US-09-345-882-4 Sequence 4, Appl |
| 7 | 20 | 0.6 | 162450 | 4 | US-09-345-882-1 Sequence 1, Appl |
| 8 | 19 | 0.6 | 45 | 3 | US-09-109-063-26 Sequence 26, Appl |
| 9 | 19 | 0.6 | 50 | 3 | US-09-109-063-2 Sequence 27, Appl |
| 10 | 19 | 0.6 | 993 | 3 | US-08-416-478A-4 Sequence 4, Appl |
| 11 | 19 | 0.6 | 999 | 1 | US-08-474-988B-4 Sequence 4, Appl |
| 12 | 19 | 0.6 | 999 | 2 | US-08-394-442B-4 Sequence 4, Appl |
| 13 | 19 | 0.6 | 999 | 2 | US-08-474-988B-4 Sequence 4, Appl |
| 14 | 19 | 0.6 | 1050 | 5 | PCR-US91-04274A-3 Sequence 3, Appl |
| 15 | 19 | 0.6 | 1069 | 5 | PCR-US91-04274A-3 Sequence 3, Appl |
| 16 | 19 | 0.6 | 1245 | 1 | US-07-887-072B-1 Sequence 1, Appl |
| 17 | 19 | 0.6 | 1245 | 1 | US-08-466-444-1 Sequence 1, Appl |
| 18 | 19 | 0.6 | 1320 | 4 | US-09-210-748A-1 Sequence 1, Appl |
| 19 | 19 | 0.6 | 1519 | 3 | US-09-109-063-3 Sequence 3, Appl |
| 20 | 19 | 0.6 | 8532 | 3 | US-08-452-655B-1 Sequence 1, Appl |
| 21 | 19 | 0.6 | 8532 | 3 | US-08-450-582-1 Sequence 1, Appl |
| 22 | 19 | 0.6 | 9606 | 1 | US-07-741-940-1 Sequence 1, Appl |
| 23 | 19 | 0.6 | 9606 | 1 | US-08-289-548A-1 Sequence 1, Appl |
| 24 | 19 | 0.6 | 9606 | 1 | US-08-452-654-1 Sequence 1, Appl |
| 25 | 19 | 0.6 | 9606 | 2 | US-08-370-235A-1 Sequence 1, Appl |
| 26 | 19 | 0.6 | 9606 | 4 | US-08-449-731-1 Sequence 1, Appl |
| 27 | 18 | 0.5 | 24 | 4 | US-08-687-580B-14 Sequence 14, Appl |

| | | | | | |
|----|----|-----|------|---|--|
| 28 | 18 | 0.5 | 301 | 4 | US-09-602-877A-81 Sequence 81, Appl |
| 29 | 18 | 0.5 | 461 | 4 | US-09-060-756-49 Sequence 49, Appl |
| 30 | 18 | 0.5 | 633 | 4 | US-09-134-001C-561 Sequence 561, Appl |
| 31 | 18 | 0.5 | 681 | 4 | US-09-134-001C-2719 Sequence 2719, Appl |
| 32 | 18 | 0.5 | 740 | 2 | US-08-713-000-8 Sequence 8, Appl |
| 33 | 18 | 0.5 | 740 | 2 | US-08-975-316-8 Sequence 8, Appl |
| 34 | 18 | 0.5 | 740 | 4 | US-09-211-710-8 Sequence 8, Appl |
| 35 | 18 | 0.5 | 740 | 4 | US-09-615-192A-8 Sequence 8, Appl |
| 36 | 18 | 0.5 | 741 | 2 | US-08-975-316-58 Sequence 58, Appl |
| 37 | 18 | 0.5 | 741 | 4 | US-09-615-192A-58 Sequence 58, Appl |
| 38 | 18 | 0.5 | 788 | 4 | US-09-615-192A-96 Sequence 96, Appl |
| 39 | 18 | 0.5 | 1139 | 1 | US-08-458-084-2 Sequence 2, Appl |
| 40 | 18 | 0.5 | 1139 | 1 | US-08-205-508-2 Sequence 2, Appl |
| 41 | 18 | 0.5 | 1139 | 1 | US-08-278-630A-11 Sequence 11, Appl |
| 42 | 18 | 0.5 | 1139 | 2 | US-08-482-148-7 Sequence 7, Appl |
| 43 | 18 | 0.5 | 1139 | 5 | PCT-US95-02944-7 Sequence 7, Appl |
| 44 | 18 | 0.5 | 1139 | 5 | PCT-US95-02945-2 Sequence 2, Appl |
| 45 | 18 | 0.5 | 2368 | 3 | US-08-714-918-2 Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-08-143-576-7
Sequence 7, Application US/08143576
Patent No. 5643761
GENERAL INFORMATION:
APPLICANT: Fisher, Paul B.
INVENTOR: Jiang, Hongping
TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRUCTURED
TITLE OF INVENTION: CNNA LIBRARY AND USES OF THE GENERATED LIBRARY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White, c/o Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,576
FILING DATE: 25-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43563/JPW/ANC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-143-576-7
Query Match 3.88; Score 129; DB 1; Length 301;
Best Local Similarity 99.48; Pred. No. 4.2e-51;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1592 ACAACGAGTATTCCTCTCTCAGATACGAGGACATACAGCTTACCTGCTTACG 1651

|||||
Db 1 ACAACGAGTATCCCTCCAGATCTGGAGCTAACAGTTCACCTGATGTTGAG 60
QY 1652 GGGCCAGAGCAAGCAAGTGAAGACATTTAAACATGTG /ATCTTGAG 1711
|||||
Db 61 GGGCCAGAGCAAGCAAGTGAAGACATTTAAACATGTG /ATCTTGAG 120
QY 1712 CATTTACTATTAACCTGTTAAAGAAACCTTGATCACTGAAACCA /TACAGGAGC 1771
|||||
Db 121 CATTACTATTAACCTGTTAAAGAAACCTTGATCACTGAAACCA /TACAGGAGC 180
|||||
RESULT 2
US-08-143-576-6
; Sequence 6, Application US/08143576
; Patent No. 5643761
; GENERAL INFORMATION:
; APPLICANT: Jiaq, Hongping
; TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRATE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, C/O Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,576
; FILING DATE: 25-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 43563/JPM/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-143-576-6
Query Match 3.2%; Score 108; DB 1; Length 193;
Best Local Similarity 100.0%; Ptd. No. 3.3e-41;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Patent No. 6444465
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wrait
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
; FILE REFERENCE: R15-0138
; CURRENT FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 182
; SEQ ID NO 24
; LENGTH: 169998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1208)...(1472)
; NAME/KEY: intron
; LOCATION: (1473)...(124390)
; NAME/KEY: exon
; LOCATION: (124391)...(124544)
; NAME/KEY: intron
; LOCATION: (124545)...(125409)
; NAME/KEY: exon
; LOCATION: (125410)...(125595)
; NAME/KEY: intron
; LOCATION: (125596)...(128711)
; NAME/KEY: exon
; LOCATION: (128712)...(128848)
; NAME/KEY: intron
; LOCATION: (128849)...(133400)
; NAME/KEY: exon
; LOCATION: (133401)...(133469)
; NAME/KEY: intron
; LOCATION: (133470)...(134652)
; NAME/KEY: exon
; LOCATION: (134653)...(134773)
; NAME/KEY: intron
; LOCATION: (134774)...(136116)
; NAME/KEY: exon
; LOCATION: (136117)...(136261)
; NAME/KEY: intron
; LOCATION: (136262)...(137936)
; NAME/KEY: exon
; LOCATION: (137937)...(138053)
; NAME/KEY: intron
; LOCATION: (138054)...(138637)
; NAME/KEY: exon
; LOCATION: (138638)...(138766)
; NAME/KEY: intron
; LOCATION: (138767)...(138864)
; NAME/KEY: exon
; LOCATION: (138865)...(138940)
; NAME/KEY: intron
; LOCATION: (138941)...(139765)
; NAME/KEY: exon
; LOCATION: (139766)...(139860)
; NAME/KEY: intron
; LOCATION: (139861)...(142245)
; NAME/KEY: exon
; LOCATION: (142246)...(142445)
; NAME/KEY: intron
; LOCATION: (142446)...(143605)
; NAME/KEY: exon
; LOCATION: (143606)...(143738)
; NAME/KEY: intron
; LOCATION: (143739)...(145838)
; NAME/KEY: exon
; LOCATION: (145839)...(145931)
; NAME/KEY: intron
; LOCATION: (145932)...(147385)
; NAME/KEY: exon
; LOCATION: (147386)...(147544)


```
NAME/KEY: Intron
LOCATION: (147545)...(153274)
NAME/KEY: exon
LOCATION: (153275)...(153321)
NAME/KEY: Intron
LOCATION: (153322)...(155088)
NAME/KEY: exon
LOCATION: (155089)...(155231)
NAME/KEY: Intron
LOCATION: (155232)...(156025)
NAME/KEY: exon
LOCATION: (156026)...(156151)
NAME/KEY: Intron
LOCATION: (156152)...(156826)
NAME/KEY: exon
LOCATION: (156827)...(156928)
NAME/KEY: Intron
LOCATION: (156929)...(163399)
NAME/KEY: exon
LOCATION: (163400)...(163586)
US-09-676-610B-24
```

```
Query Match
Best Local Similarity 100.0%; Score 22; DB 4; Length 14998;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2102 GTGATGAGGTGTGTATGA 2123
DB 149077 GTGATGAGGTGTGTATGA 149098
```

```
RESULT 4
US-09-345-882-12
Sequence 12, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 12
LENGTH: 139
TYPE: DNA
ORGANISM: Homo sapiens
US-09-345-882-12
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 1418;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2135 GTGATGAAGTGTGTATGA 2154
DB 30 GTGATGAAGTGTGTATGA 49
```

```
RESULT 5
US-09-111-470-7
Sequence 7, Application US/09111470
Patent No. 6277959
GENERAL INFORMATION:
APPLICANT: Valladeau, Jenny
APPLICANT: Ravel, Odile
APPLICANT: Bates, Elizabeth E.M.
APPLICANT: Ford, John
APPLICANT: Saeland, Sem
```

```
APPLICANT: Lebecque, Serge J.E.
TITLE OF INVENTION: Mammalian Membrane Protein Genes;
TITLE OF INVENTION: Related Reagents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,470
FILING DATE: 08-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/053,080
FILING DATE: 09-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SP0695
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 279..992
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1348
OTHER INFORMATION: /note="poly-A addition motif"
US-09-111-470-7

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 1418;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1924 GAAAAAAGCTGCAAAAA 1943
DB 510 GAAAAAAGCTGCAAAAA 529

RESULT 6
US-09-345-882-4
Sequence 4, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 4
```

| | | |
|----|---|----------------------------|
| 1 | LENGTH: 6002 | |
| 2 | TYPE: DNA | |
| 3 | ORGANISM: Homo sapiens | |
| 4 | FEATURE: | |
| 5 | NAME/KEY: allele | |
| 6 | LOCATION: 1319 | |
| 7 | OTHER INFORMATION: 5-130-257 | : polymorphic base A C T G |
| 8 | FEATURE: | |
| 9 | NAME/KEY: allele | |
| 10 | LOCATION: 1338 | |
| 11 | OTHER INFORMATION: 5-130-276 | : polymorphic base A C T G |
| 12 | FEATURE: | |
| 13 | NAME/KEY: allele | |
| 14 | LOCATION: 1944 | |
| 15 | OTHER INFORMATION: 5-136-174 | : polymorphic base C C T T |
| 16 | FEATURE: | |
| 17 | NAME/KEY: allele | |
| 18 | LOCATION: 3329 | |
| 19 | OTHER INFORMATION: 5-143-84 | : polymorphic base A C T G |
| 20 | FEATURE: | |
| 21 | NAME/KEY: allele | |
| 22 | LOCATION: 3346 | |
| 23 | OTHER INFORMATION: 5-143-101 | : polymorphic base A C T C |
| 24 | FEATURE: | |
| 25 | NAME/KEY: allele | |
| 26 | LOCATION: 4582 | |
| 27 | OTHER INFORMATION: 5-148-352 | : polymorphic base G C T T |
| 28 | FEATURE: | |
| 29 | NAME/KEY: allele | |
| 30 | LOCATION: 1107.1125 | |
| 31 | OTHER INFORMATION: polymorphic fragment 5-129-144 | SEQ T C T C |
| 32 | FEATURE: | |
| 33 | NAME/KEY: allele | |
| 34 | LOCATION: 1107.1125 | |
| 35 | OTHER INFORMATION: polymorphic fragment 5-130-276 | SEQ T C T C |
| 36 | FEATURE: | |
| 37 | NAME/KEY: allele | |
| 38 | LOCATION: 1107.1125 | |
| 39 | OTHER INFORMATION: polymorphic fragment 5-129-144 | SEQ T C T C |
| 40 | FEATURE: | |
| 41 | NAME/KEY: allele | |
| 42 | LOCATION: 1315.1338 | |
| 43 | OTHER INFORMATION: polymorphic fragment 5-130-276 | SEQ T C T C |
| 44 | FEATURE: | |
| 45 | NAME/KEY: allele | |
| 46 | LOCATION: 1921.1967 | |
| 47 | OTHER INFORMATION: polymorphic fragment 5-136-174 | SEQ T C T C |
| 48 | FEATURE: | |
| 49 | NAME/KEY: allele | |
| 50 | LOCATION: 1921.1967 | |
| 51 | OTHER INFORMATION: polymorphic fragment 5-136-174 | SEQ T C T C |
| 52 | FEATURE: | |
| 53 | NAME/KEY: allele | |
| 54 | LOCATION: 3306.3352 | |
| 55 | OTHER INFORMATION: polymorphic fragment 5-143-84 | SEQ T C T C |
| 56 | FEATURE: | |
| 57 | NAME/KEY: allele | |
| 58 | LOCATION: 3306.3352 | |
| 59 | OTHER INFORMATION: polymorphic fragment 5-143-84 | SEQ T C T C |
| 60 | FEATURE: | |
| 61 | NAME/KEY: allele | |
| 62 | LOCATION: 1296.1338 | |
| 63 | OTHER INFORMATION: polymorphic fragment 5-130-257 | SEQ T C T C |
| 64 | FEATURE: | |
| 65 | NAME/KEY: allele | |
| 66 | LOCATION: 1296.1338 | |
| 67 | OTHER INFORMATION: polymorphic fragment 5-130-257 | SEQ T C T C |
| 68 | FEATURE: | |
| 69 | NAME/KEY: allele | |
| 70 | LOCATION: 3323.3369 | |
| 71 | OTHER INFORMATION: polymorphic fragment 5-143-101 | SEQ T C T C |
| 72 | FEATURE: | |
| 73 | NAME/KEY: allele | |

| | |
|--|--|
| LOCATION: 3323..3369 | OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID65 |
| FEATURE: | |
| NAME/KEY: allele | |
| LOCATION: 4559..4605 | |
| OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID44 | |
| FEATURE: | |
| NAME/KEY: allele | |
| LOCATION: 4559..4605 | |
| OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID65 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 442..444 | |
| OTHER INFORMATION: ATG | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 4378..4380 | |
| OTHER INFORMATION: stop : TGA | |
| FEATURE: | |
| NAME/KEY: polyA_signal | |
| LOCATION: 4878..4883 | |
| OTHER INFORMATION: potential | |
| FEATURE: | |
| NAME/KEY: polyA_signal | |
| LOCATION: 5116..5121 | |
| OTHER INFORMATION: potential | |
| FEATURE: | |
| NAME/KEY: polyA_signal | |
| LOCATION: 5896..5901 | |
| OTHER INFORMATION: potential | |
| FEATURE: | |
| NAME/KEY: polyA_signal | |
| LOCATION: 5981..5986 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 209..756 | |
| OTHER INFORMATION: homology with EST in ref emb1:W645.1 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 391..815 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:W37603 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 453..898 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:H39516 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 818..1306 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:W67770 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 844..1303 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:AA262427 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 1351..1702 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:AA485189 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 1866..2109 | |
| OTHER INFORMATION: homology with EST in ref emb1:AA296993 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 2181..2281 | |
| OTHER INFORMATION: homology with EST in ref emb1:T61718 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 2253..2482 | |
| OTHER INFORMATION: homology with EST in ref emb1:AA082927 | |
| FEATURE: | |
| NAME/KEY: misc_feature | |
| LOCATION: 2480..2842 | |
| OTHER INFORMATION: complement homology with EST in ref emb1:H38607 | |

```
FEATURE:
NAME/KEY: misc-feature
LOCATION: 3334..3733
OTHER INFORMATION: homology with EST in refembl:A179595
FEATURE:
NAME/KEY: misc-feature
LOCATION: 3631..3870
OTHER INFORMATION: complement homology with EST in refembl:A169631
FEATURE:
NAME/KEY: misc-feature
LOCATION: 3683..4221
OTHER INFORMATION: homology with EST in refembl:H112
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4277..4796
OTHER INFORMATION: homology with EST in refembl:A199016
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4516..5016
OTHER INFORMATION: homology with EST in refembl:A179433
FEATURE:
NAME/KEY: misc-feature
LOCATION: 5580..6002
OTHER INFORMATION: complement homology with EST in refembl:A167428
US-09-345-882-4

Query Match      0.68; Score 20; DB 4; Length 100;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2135 GTGATGAGATGAGATGAT 2154
Db 917 GTGATGAGATGAGATGAT 936

RESULT 7
US-09-345-882-1
: Sequence 1, Application US/09345882
: Patent No. 6399373
: GENERAL INFORMATION:
: APPLICANT: Bouquelere, Lydie
: TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOLIC ACID BINDING PROTEIN (RBP-7)
: FILE REFERENCE: GENSET 031A
: CURRENT APPLICATION NUMBER: US/09/345,882
: CURRENT FILING DATE: 1999-06-30
: PRIOR APPLICATION NUMBER: US 60/091,315
: PRIOR FILING DATE: 1998-06-30
: PRIOR APPLICATION NUMBER: US 60/111,909
: PRIOR FILING DATE: 1998-12-10
: NUMBER OF SEQ ID NOS: 140
: SOFTWARE: Patent.pm
: SEQ ID NO 1
: LENGTH: 162450
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: allele
: LOCATION: 72794
: OTHER INFORMATION: 5-124-273 : polymorphic base A : G
: FEATURE:
: NAME/KEY: allele
: LOCATION: 88073
: OTHER INFORMATION: 5-127-261 : polymorphic base A : C
: FEATURE:
: NAME/KEY: allele
: LOCATION: 90842
: OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 93714
: OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
: FEATURE:

NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99096
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
: FEATURE:
: NAME/KEY: allele
```

```

LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID# 1
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID# 1
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID# 2
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-147-325 SEQ ID# 49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-147-325 SEQ ID# 70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID# 7050
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-275 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-275 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-305 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-305 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID# 71
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID# 71

```

```

OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID# 58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID# 38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID# 59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID# 39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID# 60
FEATURE:
Query Match 0.68; Score 20; DB 4; Length 162450;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 2135 GTGATGAGATGAGATGAT 2154
DB 87931 GTGATGAGATGAGATGAT 87950
RESULT 8
US-09-109-063-26
Sequence 26, Application US/09109063
Patent No. 6013498
GENERAL INFORMATION:
APPLICANT: YOKOYAMA, KEIICHI
APPLICANT: NAKAMURA, NAMI
APPLICANT: MINA, TETSUYA
APPLICANT: SEGURO, KATSUYA
TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 0010-0937-0
CURRENT APPLICATION NUMBER: US/09/109, 063
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: JP 180010/1997
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 45
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC DNA
US-09-109-063-26
Query Match 0.68; Score 19; DB 3; Length 45;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 490 GAGACGCTCATGATGAT 508
DB 10 GAGACGCTCATGATGAT 28
RESULT 9
US-09-109-063-27/c
Sequence 27, Application US/09109063
Patent No. 6013498
GENERAL INFORMATION:
APPLICANT: YOKOYAMA, KEIICHI
APPLICANT: NAKAMURA, NAMI
APPLICANT: MINA, TETSUYA
APPLICANT: SEGURO, KATSUYA
TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
FILE REFERENCE: 0010-0937-0

```

;; CURRENT APPLICATION NUMBER: US/09/109,063
;; CURRENT FILING DATE: 1998-07-02
;; EARLIER APPLICATION NUMBER: JP 180010/1997
;; EARLIER FILING DATE: 1997-07-04
;; NUMBER OF SEQ ID NOS: 62
;; SOFTWARE: PatentIn Ver. 2.0
;; LENGTH: 50
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC DNA
US-09-109-063-27
Query Match 0.6%; Score 19; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 490 GAGACGCTCATGATGAAT 508
DB 45 GAGACGCTCATGATGAAT 27
RESULT 10
US-09-109-063-2
;; Sequence 2, Application US/09109063
;; Patent No. 6013498
;; GENERAL INFORMATION:
;; APPLICANT: YOKOTAMA, KEIICHI
;; APPLICANT: NAKAMURA, NAMI
;; APPLICANT: MIWA, TETSUYA
;; APPLICANT: SEGURO, KATSUYA
;; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
;; FILE REFERENCE: 0010-0937-0
;; CURRENT APPLICATION NUMBER: US/09/109,063
;; CURRENT FILING DATE: 1998-07-02
;; EARLIER APPLICATION NUMBER: JP 180010/1997
;; EARLIER FILING DATE: 1997-07-04
;; NUMBER OF SEQ ID NOS: 62
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 2
;; LENGTH: 993
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC DNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(993)
;; OTHER INFORMATION: IDENTIFICATION METHOD: S
US-09-109-063-2
Query Match 0.6%; Score 19; DB 3; Length 993;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 490 GAGACGCTCATGATGAAT 508
DB 412 GAGACGCTCATGATGAAT 430
RESULT 11
US-08-416-478A-4/C
;; Sequence 4, Application US/08416478A
;; Patent No. 5773578
;; GENERAL INFORMATION:
;; APPLICANT: Hercend, Thierry
;; APPLICANT: Triebel, Frederic
;; TITLE OF INVENTION: New Proteins Produced By Human
;; TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
;; TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;; STREET: 419 Seventh Street, N.W., Suite 400
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/416,478A
;; FILING DATE: 04-APR-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/854,644
;; FILING DATE: 08-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: YUN, Allen C.
;; REGISTRATION NUMBER: 37,971
;; REFERENCE/DOCKET NUMBER: HERCEND-1A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 628-5197
;; TELEFAX: (202) 737-3528
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 999 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-416-478A-4
Query Match 0.6%; Score 19; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 880 TGGGGCATGAGAGATRACT 898
DB 140 TGGGGCATGAGAGATRACT 122
RESULT 12
US-08-474-988B-4/C
;; Sequence 4, Application US/08474988B
;; Patent No. 5874250
;; GENERAL INFORMATION:
;; APPLICANT: Hercend, Thierry
;; APPLICANT: Triebel, Frederic
;; TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN
;; TITLE OF INVENTION: LYMPHOCYTES, DNA SEQUENCES ENCODING THESE PROTEINS AND
;; TITLE OF INVENTION: THEIR PHARMACEUTICAL AND BIOLOGICAL USES
;; NUMBER OF SEQUENCES: 11
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;; STREET: 419 Seventh Street, N.W., Suite 400
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/474,988B
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/416,478
;; FILING DATE: 04-APR-1995
;; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,644
FILING DATE: 08-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: HERCEND-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-474-988B-4

Query Match 0.6%; Score 19; DB 2; Length 1000;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 880 TGGGCGATGAGATAACT 898
|||||
DB 140 TGGGCGATGAGATAACT 122

RESULT 13
US-08-394-442B-4/C
Sequence 4, Application US/08394442B
Patent No. 5976877
GENERAL INFORMATION:
APPLICANT: Hercend, Thierry
REFERENCE/DOCKET NUMBER: 32585
TITLE OF INVENTION: New Proteins Produced by Human
TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,442B
FILING DATE: 24-FEB-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HERCEND-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-394-442B-4

Query Match 0.6%; Score 19; DB 2; Length 1000;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 880 TGGGCGATGAGATAACT 898
|||||
DB 140 TGGGCGATGAGATAACT 122

RESULT 14
PCT-US91-04274A-4
Sequence 4, Application PC/TUS9104274A
GENERAL INFORMATION:
APPLICANT: Williams, Douglas E.
APPLICANT: Lyman, Stewart
TITLE OF INVENTION: Mast Cell Growth Factor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/04274A
FILING DATE: 19910614
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oster, Jeffrey B.
REGISTRATION NUMBER: 32585
REFERENCE/DOCKET NUMBER: 0521D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ORIGINAL SOURCE:
ORGANISM: HUMAN MAST CELL GROWTH FACTOR
IMMEDIATE SOURCE:
CLONE: HMGF-2.4
PCT-US91-04274A-4

Query Match 0.6%; Score 19; DB 5; Length 1050;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1214 AGAAGAAAAAGCATCTGA 1232
|||||
DB 574 AGAAGAAAAAGCATCTGA 592

RESULT 15
PCT-US91-04274A-3
Sequence 3, Application PC/TUS9104274A
GENERAL INFORMATION:
APPLICANT: Williams, Douglas E.
APPLICANT: Lyman, Stewart
TITLE OF INVENTION: Mast Cell Growth Factor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US91/04274A
 FILING DATE: 19910614
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Oster, Jeffrey B.
 REGISTRATION NUMBER: 32585
 REFERENCE/DOCKET NUMBER: 0521D
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 2065870430
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1069 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: N
 ANTI-SENSE: N
 ORIGINAL SOURCE:
 ORGANISM: HUMAN MAST CELL GROWTH FACTOR
 IMMEDIATE SOURCE:
 CLONE: HMGF-2D
 PCT-US91-04274A-3

Query Match

0.6%; Score 19; DB 5; Length 1159;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1214 AGAAGAAAAAGCATCTGA 1232
 ||||||||||||||||
 DB 575 AGAAGAAAAAGCATCTGA 593

Search completed: May 16, 2003, 20:08:09
 Job time : 1831 secs

PN MO200164707-A1.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US06960.
 XX 29-FEB-2000; 2000US-0515363.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 XX PI Fisher PB, Kang D, GopalKrishnan RV;
 XX DR MPI: 2001-565494/63.
 XX P-PSDB: AAEI0155.
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity -
 XX
 PS Claim 1; Page 16-18; 152pp; English.
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC antileukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human Mda-5 cDNA.
 XX
 SO Sequence 3365 BP; 1138 A; 644 C; 753 G; 830 T; 0 other:
 Query Match 100.0%; Score 3365; DB 22; Length 3365;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 421 AGACCGGCGACCCCTCTGGCCGCCGCTACATGAACTTGAGCTCAGGACTTCCCTCT 480
 QY 481 CCATGTTTGAGAGAGCTCATGATGATATATCCAACTGTGAACCTCTTCAAGCCACT 540
 Db 481 CCATGTTTGAGAGAGCTCATGATGATATATCCAACTGTGAACCTCTTCAAGCCACT 540
 QY 541 CTGTGGACAGCTTCTAGTTAGAGAGCTTTGGATTAAAGTCATGAGAGAGAACTGTTG 600
 Db 541 CTGTGGACAGCTTCTAGTTAGAGAGCTTTGGATTAAAGTCATGAGAGAGAACTGTTG 600
 QY 601 ACAATTGAG 660
 Db 601 ACAATTGAG 660
 QY 661 AGAGAGCTACTAAAAAGATTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 Db 661 AGAGAGCTACTAAAAAGATTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 721 CTTGCTCAAG 780
 Db 721 CTTGCTCAAG 780
 QY 781 AGCAATGAG 840
 Db 781 AGCAATGAG 840
 QY 841 CTTTCAACCCAGTTCAG 900
 Db 841 CTTTCAACCCAGTTCAG 900
 QY 901 TCAGATATCTTTTGGAG 960
 Db 901 TCAGATATCTTTTGGAG 960
 QY 961 GGAATGTCAG 1020
 Db 961 GGAATGTCAG 1020
 QY 1021 GGCACCATGAG 1080
 Db 1021 GGCACCATGAG 1080
 QY 1081 GAACTCAG 1140
 Db 1081 GAACTCAG 1140
 QY 1141 ATCATCATCTGCTCCCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Db 1141 ATCATCATCTGCTCCCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1201 GATCAGTTAG 1260
 Db 1201 GATCAGTTAG 1260
 QY 1261 AAGGTACGTGATGAG 1320
 Db 1261 AAGGTACGTGATGAG 1320
 QY 1321 TATGCTGTTATGATTAAGTGGTATACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 Db 1321 TATGCTGTTATGATTAAGTGGTATACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1381 AAGTCTGTGAT 1440
 Db 1381 AAGTCTGTGAT 1440
 QY 1441 GAAATGAG 1500
 Db 1441 GAAATGAG 1500
 QY 1501 TGTATACACCAAG 1560
 Db 1501 TGTATACACCAAG 1560

| QY | 2641 | GCTGATGAGACACTACGTCCTGTTGGTCACAGTGGTTGAGAGTATCGAAATGAG | 270 |
|----------|--|---|-----|
| Db | 2641 | GGTGTGTGAGACACCCACGTCCTGTTGGTCACAGTGGTTGAGAGTATCGAAATGAG <th>270</th> | 270 |
| QY | 2701 | ACAGTTAATGATTTCCGAGAGAGATGATGATTAAGCTATACATTTGTGTCAAAATATG <th>276</th> | 276 |
| Db | 2701 | ACAGTTAATGATTTCCGAGAGAGATGATGATTAAGCTATACATTTGTGTCAAAATATG <th>276</th> | 276 |
| QY | 2761 | AAACCCAGAGAGATGCTCATAGATTTTGGAAATTACAGATGCAAGATATATGGFAAG <th>282</th> | 282 |
| Db | 2761 | AAACCCAGAGAGATGCTCATAGATTTTGGAAATTACAGATGCAAGATATATGGFAAG <th>282</th> | 282 |
| QY | 2821 | AAATGAAAAACCAAGAGAAATATTTGCCAGAGATTACAAAGATTAACCATCACTAACT <th>288</th> | 288 |
| Db | 2821 | AAATGAAAAACCAAGAGAAATATTTGCCAGAGATTACAAAGATTAACCATCACTAACT <th>288</th> | 288 |
| QY | 2881 | TTCCCTTCCAAAACTGACAGTGTGCTAGCCGTCTTGCGGAGATATCCATGATATAG <th>294</th> | 294 |
| Db | 2881 | TTCCCTTCCAAAACTGACAGTGTGCTAGCCGTCTTGCGGAGATATCCATGATATAG <th>294</th> | 294 |
| QY | 2941 | AAATGATCAGCTCATATATGACCCGAGAAATTCAGAGAACTTACATTTGTAAGAAAC <th>300</th> | 300 |
| Db | 2941 | AAATGATCAGCTCATATATGACCCGAGAAATTCAGAGAACTTACATTTGTAAGAAAC <th>300</th> | 300 |
| QY | 3001 | AAAGCATGTCAGCAAAAGAAATGTGCCGACATATCAATTAATGTTGAATCATCTG <th>306</th> | 306 |
| Db | 3001 | AAAGCATGTCAGCAAAAGAAATGTGCCGACATATCAATTAATGTTGAATCATCTG <th>306</th> | 306 |
| QY | 3061 | GGCCAGGCTTGGGGAGACATATGCTGACCAAGGCTTAGATTGGCTGTGTCAAAATA <th>312</th> | 312 |
| Db | 3061 | GGCCAGGCTTGGGGAGACATATGCTGACCAAGGCTTAGATTGGCTGTGTCAAAATA <th>312</th> | 312 |
| QY | 3121 | AGGAAATTTGTAGTGGTTTCAAAAATATATTCACAAAGAAACAAATCAAAAAGTGGTA <th>318</th> | 318 |
| Db | 3121 | AGGAAATTTGTAGTGGTTTCAAAAATATATTCACAAAGAAACAAATCAAAAAGTGGTA <th>318</th> | 318 |
| QY | 3181 | GAATTACCTATACATTTTCCCAATCTTGACATTCACAGATGCTTTATTAATGATAG <th>324</th> | 324 |
| Db | 3181 | GAATTACCTATACATTTTCCCAATCTTGACATTCACAGATGCTTTATTAATGATAG <th>324</th> | 324 |
| QY | 3241 | GATTGACCTGATGAGATTTCTTTAAATACATGATGATTAACATTTAAATGATTA <th>330</th> | 330 |
| Db | 3241 | GATTGACCTGATGAGATTTCTTTAAATACATGATGATTAACATTTAAATGATTA <th>330</th> | 330 |
| QY | 3301 | TGATTAAATGATTTATGCTAGACAGAACTGACATTAAGAAATCAATAAATGATTTTA <th>336</th> | 336 |
| Db | 3301 | TGATTAAATGATTTATGCTAGACAGAACTGACATTAAGAAATCAATAAATGATTTTA <th>336</th> | 336 |
| QY | 3361 | CTCTG 3365 <th></th> | |
| Db | 3361 | CTCTG 3365 <th></th> | |
| RESULT 2 | | | |
| ID | ABA04908 | standard; cDNA: 3372 BP. | |
| AC | ABA04908; | | |
| XX | 05-MAR-2002 | (first entry) | |
| XX | Human RNA helicase RH16 coding sequence. | | |
| XX | Human: RH16; RNA helicase; cytosolic; vifucide; anti-HIV; | | |
| KW | immunosuppressive; immunostimulatory; antipneumatic; anti-influc; | | |
| KW | antiarthritis; osteopathic; antidiabetic; hepatotrophic; | | |
| KW | antimicrobial; cancer; infection; HIV; hepatitis; genetic disease; | | |
| KW | autoimmune disease; graft rejection; vaccine; ss. | | |
| XX | Homo sapiens. | | |
| XX | Key | Location/Qualifiers | |
| XX | CDS | 155..3232 | |
| XX | LT | | |

```

FT      /*tag- a
FT      /product- "Human RH116"
PN      WO200185955-A1.
XX      15-NOV-2001.
XX      11-MAY-2001: 2001WO-FR01441.
XX      11-MAY-2000: 2000FR-0006030.
XX      (ISTA-) ISTAC.
XX      (INSP) INST PASTEUR LILLE.
XX      Bahr G, Cocude C, Capron A;
XX      WPI: 2002-082898/11.
XX      P-PSDB: AAM47798.
XX
PT      New polypeptide, useful for treating and diagnosing cancer or
PT      inflammation, and drug screening, comprises a human polynucleotide
PT      homologous to RNA helicase
XX
PS      Claim 7: Page 85-89; 114pp: French.
XX
CC      The present sequence is the coding sequence for human RH116. RH116 is a
CC      116kDa protein and has homology to RNA helicases (DEXH box). RH116 and
CC      its coding sequence are useful for treating cancer; acute or chronic
CC      infections (especially by HIV or hepatitis B or C); inherited genetic
CC      diseases (autoimmune diseases (particularly rheumatism, arthritis,
CC      arteriosclerosis, osteoporosis and diabetes, but many others listed) and
CC      to prevent graft rejection. RH116 and its coding sequence are also useful
CC      for inducing, or increasing, the immune response to a vaccine.
XX
SQ      Sequence 3372 BP; 1157 A; 637 C; 748 G; 830 T; 0 other:
Query Match      90.4%; Score 3042; DB 24; Length 3372;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3342; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db      483 CTCATGATGAATATCTCCAACTGCTGAACCTCTTCAACCCACCTGCTGAGCAGCTTG 542
Oy      TAGTAGAGAGAGCTCTTGATTAAGTCATGAGAGAGAACTGTGACATTTGACACAGAA 616
Db      543 TAGTTAGAGAGAGCTCTTGATTAAGTCATGAGAGAGAACTGTGACATTTGACACAGAA 602
Oy      617 ACCGAGATTGCTGCTGACAGAAAACATGAAATGAAATCAGTGTAAAGAGCTACTAAAAA 676
Db      603 ACCGAGATTGCTGCTGACAGAAAACATGAAATGAAATCAGTGTAAAGAGCTACTAAAAA 662
Oy      677 GGATTGTGCAAGAAAGAAAGCTGTCTGCTGATTTGAAATGCTTCTGCTCAACAGGAA 736
Db      663 GGATTGTGCAAGAAAGAAAGCTGTCTGCTGATTTGAAATGCTTCTGCTCAACAGGAA 722
Oy      737 ACAATGAACCTGTCCAAAGAGTTAAACAGAGCTGTGATTTGCTGAGAAAGCAATGACAGATTG 796
Db      723 ACAATGAACCTGTCCAAAGAGTTAAACAGAGCTGTGATTTGCTGAGAAAGCAATGACAGATTG 782
Oy      797 AGAATTTATCAAGAGTTGATGCTGCTCAAGTGAAGAGCAACCTTTTCAACACAGCTTC 856
Db      783 AGAATTTATCAAGAGTTGATGCTGCTCAAGTGAAGAGCAACCTTTTCAACACAGCTTC 842
Oy      857 AGCCAAATCTGGAAGAGAGGCTGTGGGCTATGAGAAATCACTCAAGATCATCTTTTG 916
Db      843 AGCCAAATCTGGAAGAGAGGCTGTGGGCTATGAGAAATCACTCAAGATCATCTTTTG 902
Oy      917 CAGATTTCTTGATGATTTTCAAGATCAACAGAGCTTTGGAGAAAGAAAGTGTACAGTGC 976
Db      903 CAGATTTCTTGATGATTTTCAAGATCAACAGAGCTTTGGAGAAAGAAAGTGTACAGTGC 962
Oy      977 TAGATGAAGCTTGTGACATTAACAGCAACATGAGGAGTATTCAGGACCATGAGAAAGTG 1036
Db      963 TAGATGAAGCTTGTGACATTAACAGCAACATGAGGAGTATTCAGGACCATGAGAAAGTG 1022
Oy      1037 ATTGATGAAGAGAAATGTGACAGAAAGAGATCCCGGAGCCAGCACTCCAGCTGAG 1096
Db      1023 ATTGATGAAGAGAAATGTGACAGAAAGAGATCCCGGAGCCAGCACTCCAGCTGAG 1082
Oy      1097 CTTACCAAAATGGAAGTTGCTCCAGCAGAGCTTTGGAAGAGGAAAGAAATATCATCTGCTCC 1156
Db      1083 CTTACCAAAATGGAAGTTGCTCCAGCAGAGCTTTGGAAGAGGAAAGAAATATCATCTGCTCC 1142
Oy      1157 CTACAGGAGTGTGAAAAACAGAGTGTGTTTACATTTGCCAAGATCATTACACAGAA 1216
Db      1143 CTACAGGAGTGTGAAAAACAGAGTGTGTTTACATTTGCCAAGATCATTACACAGAA 1202
Oy      1217 AGAAAAAGCATGTGAGGCTGTGAAAAAGTTATGATTTGTCGAATAGGATCTGCTAGTTG 1276
Db      1203 AGAAAAAGCATGTGAGGCTGTGAAAAAGTTATGATTTGTCGAATAGGATCTGCTAGTTG 1262
Oy      1277 AACACCTCTTCCGCAAGAGAGTTCCAAACATTTTGAAGAAATGTATGCTTTATGGAT 1336
Db      1263 AACACCTCTTCCGCAAGAGAGTTCCAAACATTTTGAAGAAATGTATGCTTTATGGAT 1322
Oy      1337 TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTTGTCAAGTCTGTATATTA 1396
Db      1323 TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTTGTCAAGTCTGTATATTA 1382
Oy      1397 TTATGATGAGAGCTCAAAATGCTTGAAGAAAGCTGCTTTAACTTGAAGAAATGAGAAAGT 1456
Db      1383 TTATGATGAGAGCTCAAAATGCTTGAAGAAAGCTGCTTTAACTTGAAGAAATGAGAAAGT 1442
Oy      1457 CTGCTGTCAATTTGTCAAGCTTTTCCCTCATATCATATGATGAATGTCTATCAACACAA 1516
Db      1443 CTGCTGTCAATTTGTCAAGCTTTTCCCTCATATCATATGATGAATGTCTATCAACACAA 1502
Oy      1517 AAGAAGCAGTGTATATATATATCATATGAGGCAATTTATGATGACAGAAAGTGAAGAAACATA 1576
Db      1503 AAGAAGCAGTGTATATATATATATCATATGAGGCAATTTATGATGACAGAAAGTGAAGAAACATA 1562
Oy      1577 GACTCAAGAAAGAAACAAACAGATGATGCTTCTCCACATGATGAGATACAGCTT 1636

```


XX Fisher PB, Kang D, GopalKrishnan RV;
 XX WPI; 2001-565494/63.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity
 XX
 PS Disclosure: Page 134-148; 152pp; English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC Note: The present sequence is human Mda-5 protein-related DNA.
 CC The present sequence is designated SEQ ID NO:2 in the sequence
 CC listing, but does not correspond to the sequence designated SEQ ID NO:2
 CC in the main body of the specification (AAE10155). The present sequence is
 CC not further referred to in the specification, and has been represented
 CC in a protein format in the sequence listing.
 XX
 SQ Sequence 3131 BP; 1075 A; 583 C; 699 G; 774 T; 0 other;
 Query Match 80.5%; Score 2709; DB 22; Length 3131;
 Best Local Similarity 99.8%; Pred No. 0;
 Matches 3129; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 709 TTTCTGAATGTTCTTCTGTCACAAACAGAAACATGAACTGTCCAAAGCTTAACAGCTCT 768
 DB 541 TTTCTGAATGTTCTTCTGTCACAAACAGAAACATGAACTGTCCAAAGCTTAACAGCTCT 600
 QY 769 GATTCTCTAGAAAGCATGACAGAGATTGAGAAATTTATACAAAGTTGATGCTCTACAGT 828
 DB 601 GATTCTCTAGAAAGCATGACAGAGATTGAGAAATTTATACAAAGTTGATGCTCTACAGT 660
 QY 829 GAGAGCAACTCTTCTCAACACAGATTGAGCAATCTGGAAGAGAGCTTGAGGCTATG 888
 DB 661 GAGAGCAACTCTTCTCAACACAGATTGAGCAATCTGGAAGAGAGCTTGAGGCTATG 720
 QY 889 GAGAAATTAATCTATGAAATCATCTTTTGCAGATTTCTCTGATGTTGAGTTGAGAAACACA 948
 DB 721 GAGATTAATCTATGAAATCATCTTTTGCAGATTTCTCTGATGTTGAGTTGAGAAACACA 780
 QY 949 AGTTTGSCAGAAAGAGTGCAGCTCTTACATGAAAGCTTTGACATACAGCAATATG 1008
 DB 781 AGTTTGSCAGAAAGAGTGCAGCTCTTACATGAAAGCTTTGACATACAGCAATATG 840
 QY 1009 GGCAGTATTCAGAGCAGCATGGAAGTATTCATGATGAAGAAATGTGGCAAGCAAGCA 1068
 DB 841 GGCAGTATTCAGAGCAGCATGGAAGTATTCATGATGAAGAAATGTGGCAAGCAAGCA 900
 QY 1069 TCCCGGAGCCAGAACTCCAGCTCAGGCTTACCAAAATGGAAGTTGCCACCCACCTTG 1128
 DB 901 TCCCGGAGCCAGAACTCCAGCTCAGGCTTACCAAAATGGAAGTTGCCACCCACCTTG 960
 QY 1129 GAGGGAAGAAATATCATCATCTGCTCTCTCTACAGGAGTGGAAAAACAGAGTGTGTT 1188
 DB 961 GAGGGAAGAAATATCATCATCTGCTCTCTCTACAGGAGTGGAAAAACAGAGTGTGTT 1020
 QY 1189 TACATTCGCAAGGATCACTTAGACAAAGAAAGAAAGATCTGAGCCGGAAGAAATATA 1248
 DB 1021 TACATTCGCAAGGATCACTTAGACAAAGAAAGAAAGATCTGAGCCGGAAGAAATATA 1080
 QY 1249 GTTCTTGCAATTAAGTACTCTAGTAGTACAGCTCTTCCGCAAGAGATTCCAAATTT 1308
 DB 1081 GTTCTTGCAATTAAGTACTCTAGTAGTACAGCTCTTCCGCAAGAGATTCCAAATTT 1140
 QY 1309 TTGAAGAAATGATATCATCTGCTCTCTCTCTACAGGAGTGGAAAAACAGAGTGTGTT 1368
 DB 1141 TTGAAGAAATGATATCATCTGCTCTCTCTCTACAGGAGTGGAAAAACAGAGTGTGTT 1200
 QY 1369 CCAGAAAGTTGCAAGTCTCTGATATTATATCATGATAGCTCAAAATCTTGAAGATCTC 1428
 DB 1201 CCAGAAAGTTGCAAGTCTCTGATATTATATCATGATAGCTCAAAATCTTGAAGATCTC 1257
 QY 1429 CTCTTAACTTGAAAAATGGAAGAGATGCTGTTCAATTTGTACAGCTTTTCCATTT 1488
 DB 1258 CTCTTAACTTGAAAAATGGAAGAGATGCTGTTCAATTTGTACAGCTTTTCCATTT 1317
 QY 1489 ATCATATGATGATGTCATACACCAACAAAGACAGTGTATTAATCATGATGAGCAT 1548
 DB 1318 ATCATATGATGATGTCATACACCAACAAAGACAGTGTATTAATCATGATGAGCAT 1377
 QY 1549 TATTGATGACGAAGTTGAAAAACATATGACTCAAGAAAGAAACCAACAGATGATTC 1608
 DB 1378 TATTGATGACGAAGTTGAAAAACATATGACTCAAGAAAGAAACCAACAGATGATTC 1437
 QY 1609 CTTCCTAGATAGTGGAGCAATACAGCTTCACTGCTGTTGGAGGAGCCACAGAGTAAGCC 1668
 DB 1438 CTTCCTAGATAGTGGAGCAATACAGCTTCACTGCTGTTGGAGGAGCCACAGAGTAAGCC 1497
 QY 1669 AAAGCTGAAGAACACATTTTAAACATATGCTCCAACTCTGATGATTTACTATTAAACT 1728
 DB 1498 AAAGCTGAAGAACACATTTTAAACATATGCTCCAACTCTGATGATTTACTATTAAACT 1557
 QY 1729 GTTAAAGAAACCTTGATCACTGAAAAACCAATATACAGAGCCATCAGAAATATTCGCC 1788
 DB 1558 GTTAAAGAAACCTTGATCACTGAAAAACCAATATACAGAGCCATCAGAAATATTCGCC 1617

QY 1789 ATTGCAGATGCAACGAGAGAGATCCATTTAAAGAGAACTTGTAGAAAATATGACAAAG 1848
 DB 1618 ATTGCAGATGCAACGAGAGAGATCCATTTAAAGAGAACTTGTAGAAAATATGACAAAG 1677
 QY 1849 ATTCAAACCTTATGTCAAATGAGTCCATGATGATTTTGGAAATGCAACATATGAAACAA 1908
 DB 1678 ATTCAAACCTTATGTCAAATGAGTCCATGATGATTTTGGAAATGCAACATATGAAACAA 1737
 QY 1909 TGGGCGCATTCGAATGAG 1968
 DB 1738 TGGGCGCATTCGAATGAG 1797
 QY 1969 GCAGAACATTTGAG 2028
 DB 1798 GCAGAACATTTGAG 1857
 QY 2029 GAT--GCCTATCTCATCTTGAACCTTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAG 2086
 DB 1858 GATCGCGCTATCTCATCTTGAACCTTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAG 1917
 QY 2087 TCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2146
 DB 1918 TCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
 QY 2147 AGGATGATTTAAAG 2206
 DB 1978 AGGATGATTTAAAG 2037
 QY 2207 TTTTGGAG 2266
 DB 2038 TTTTGGAG 2097
 QY 2267 AGCTGACCAATTTAG 2326
 DB 2098 AGCTGACCAATTTAG 2157
 QY 2327 GAGGATATCTTTTACAG 2386
 DB 2158 GAGGATATCTTTTACAG 2217
 QY 2387 AAAATGAG 2446
 DB 2218 AAAATGAG 2277
 QY 2447 GAG 2506
 DB 2278 GAG 2337
 QY 2507 CTGAG 2566
 DB 2338 CTGAG 2397
 QY 2567 AATGTACATTTGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2626
 DB 2398 AATGTACATTTGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2457
 QY 2627 GTGGGTGAG 2686
 DB 2458 GTGGGTGAG 2517
 QY 2687 TTATGAG 2746
 DB 2518 TTATGAG 2577
 QY 2747 GTGTTCAG 2806
 DB 2578 GTGTTCAG 2637
 QY 2807 GTATATGAG 2866
 DB 2638 GTATATGAG 2697
 QY 2867 CATCACTATTAACCTTCTTCCAAAAAAGTGCAGTGTGCTAGCTGTTCCTGAGAGAGATA 2926

DB 2698 CATCACTATTAACCTTCTTCCAAAAAAGTGCAGTGTGCTAGCTGTTCCTGAGAGAGATA 2757
 QY 2927 TCATGATATATGAG 2986
 DB 2758 TCATGATATATGAG 2817
 QY 2987 TTGTAG 3046
 DB 2818 TTGTAG 2877
 QY 3047 TCATCTGAG 3106
 DB 2878 TCATCTGAG 2937
 QY 3107 CTGTCTCAAAATATGAG 3166
 DB 2938 CTGTCTCAAAATATGAG 2997
 QY 3167 ACAAAAGTGGGTAG 3226
 DB 2998 ACAAAAGTGGGTAG 3057
 QY 3227 TATTAGTATGAG 3286
 DB 3058 TATTAGTATGAG 3117
 QY 3287 ATTATATATGATTA 3300
 DB 3118 ATTATATATGATTA 3131

RESULT 4
 AAS40960
 ID AAS40960 standard; cDNA: 1967 BP.
 AC AAS40960;
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human enzyme polypeptide #176.
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW lipase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 OS Homo sapiens.
 PN WO20015301-A2.
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0233968.
 PR 14-SEP-2000; 2000US-0233997.
 PR 14-SEP-2000; 2000US-0233998.
 PR 14-SEP-2000; 2000US-0233999.
 PR 14-SEP-2000; 2000US-0234000.
 PR 14-SEP-2000; 2000US-0234001.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246534.
 PR 08-NOV-2000; 2000US-0246535.
 PR 08-NOV-2000; 2000US-0246536.
 PR 08-NOV-2000; 2000US-0246537.
 PR 08-NOV-2000; 2000US-0246538.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249268.
 PR 17-NOV-2000; 2000US-0249269.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250161.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251031.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251857.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-465566/50.

P-PSDB; AAU23090.

Novel polypeptides and polynucleotides useful for diagnosis, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases

Claim 4; SEQ ID No 186; 1180bp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis).

CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPRO
 CC at ftp.wipro.int/pub/published_pcl_sequences.

XX Sequence 1967 BP; 723 A; 344 C; 416 G; 476 T; 8 other:

Query Match 42.2%; Score 1419; DB 22; Length 1967;

Best Local Similarity 99.9%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;

Matches 1469; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1047 AGGAAATGTGGCAGCAAGAGCATCCCGGAGCAGCAACCTCCAGCTGAGCTTACCAAT 1106
 DB 2 AGGAAATGTGGCAGCAAGAGCATCCCGGAGCAGCAACCTCCAGCTGAGCTTACCAAT 61

QY 1107 GGAAAGTGGCCAGCCAGCTTGGAGGAGAAATATCATCTGCTCTACAGGAG 1166
 DB 62 GGAAAGTGGCCAGCCAGCTTGGAGGAGAAATATCATCTGCTCTACAGGAG 121

QY 1167 TGGAAAAACCCAGAGTGGCTGTACATTCGCAAGATCAGTACAGCAAGAAAAAGC 1226
 DB 122 TGGAAAAACCCAGAGTGGCTGTACATTCGCAAGATCAGTACAGCAAGAAAAAGC 181

QY 1227 ATCTGAGCTGGAAAAATTAAGTCTTCTCAATTAAGTACTGCTAGTAAACAGCTTT 1286
 DB 182 ATCTGAGCTGGAAAAATTAAGTCTTCTCAATTAAGTACTGCTAGTAAACAGCTTT 241

QY 1287 CCGCAAGAGTCCAGCAACCTTTTGAAGAAATGATGCTGATTAAGAAAGTGTGA 1346
 DB 242 CCGCAAGAGTCCAGCAACCTTTTGAAGAAATGATGCTGATTAAGAAAGTGTGA 301

QY 1347 TACCAAGTGAATAATATCATTTCCAGAGTGTCAAGTCTGTGATATATATCAAGTAC 1406
 DB 302 TACCAAGTGAATAATATCATTTCCAGAGTGTCAAGTCTGTGATATATATCAAGTAC 361

QY 1407 AGCTCAATCTTGAAGAACTCCCTCTTAACTTGAAGAAATGAGAGATCTGTGTGTA 1466
 DB 362 AGCTCAATCTTGAAGAACTCCCTCTTAACTTGAAGAAATGAGAGATCTGTGTGTA 421

QY 1467 ATGTGAGACTTTCCTCATTTATCATGTGTAATGTCATCACCAACAGAGAACAGT 1526
 DB 422 ATGTGAGACTTTCCTCATTTATCATGTGTAATGTCATCACCAACAGAGAACAGT 481

QY 1527 GTATTAATCATCATGAGGATATTTGATGAGAAAGTGAAGAAACATTCAGCTCAGAA 1586
 DB 482 GTATTAATCATCATGAGGATATTTGATGAGAAAGTGAAGAAACATTCAGCTCAGAA 541

QY 1587 AGAAAAACAACAGATATTCCTCTCCAGATACTGGAGATTAACAGCTTACCTGCTGT 1646
 DB 542 AGAAAAACAACAGATATTCCTCTCCAGATACTGGAGATTAACAGCTTACCTGCTGT 601

QY 1647 TGGAGGGGCGCAGCAAGAGCCAAAGCTGAAGAACACATTTTAAACTAATGCGCAATCT 1706
 DB 602 TGGAGGGGCGCAGCAAGAGCCAAAGCTGAAGAACACATTTTAAACTAATGCGCAATCT 661

QY 1707 TGATGATTTACTATTAAACTGTAAAGAAACCTTGATCAATGAATACCAATATCA 1766
 DB 662 TGATGATTTACTATTAAACTGTAAAGAAACCTTGATCAATGAATACCAATATCA 721

QY 1767 GAGAGCATGCAAGAGTGGCATTGCGAGATGCAACCAAGAGATCAATTAAGAGAA 1826
 DB 722 GAGAGCATGCAAGAGTGGCATTGCGAGATGCAACCAAGAGATCAATTAAGAGAA 781

QY 1827 ACTCTAGAAATTAATGACAAGATTCAAACTTATGTCAAATAGTCCAAATGTCAGATTT 1886
 DB 782 ACTCTAGAAATTAATGACAAGATTCAAACTTATGTCAAATAGTCCAAATGTCAGATTT 841

QY 1887 TGGAACTCAACCTATGAAACATGGGCAATTCAAATGAAAAAAGCTTCAAAAAAGG 1946
 DB 1887 TGGAACTCAACCTATGAAACATGGGCAATTCAAATGAAAAAAGCTTCAAAAAAGG 1946

DB 842 TGGAACTCAACCTATGAAACATGGGCAATTCAAATGAAAAAAGCTTCAAAAAAGG 901

QY 1947 AATGCGAAAGAAAGCTGTTTGGCAGACATTTAGGAGATGCAATGAGGCCCTACCAAT 2006
 DB 902 AATGCGAAAGAAAGCTGTTTGGCAGACATTTAGGAGATGCAATGAGGCCCTACCAAT 961

QY 2007 TAATGACCAATTCGAATGATAGATGCTATACATCTTGAACCTTCTATTAATGAAGA 2066
 DB 962 TAATGACCAATTCGAATGATAGATGCTATACATCTTGAACCTTCTATTAATGAAGA 1021

QY 2067 GAAAGATTAAGAGTTTGCAGTCATTAAGAGATGATAGATGAGGTGCTATGATGACTA 2126
 DB 1022 GAAAGATTAAGAGTTTGCAGTCATTAAGAGATGATAGATGAGGTGCTATGATGACTA 1081

QY 2127 TTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2186
 DB 1082 TTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1141

QY 2187 TAGATTTTCATGACTTATTTTGGAAAAACATTAATGTTGAAAAAGCTGGCTGAAAA 2246
 DB 1142 TAGATTTTCATGACTTATTTTGGAAAAACATTAATGTTGAAAAAGCTGGCTGAAAA 1201

QY 2247 CCGAGAAATATGAATAATGAAGCTGACCAATTAAGAAATACCATTAATGAGCAATATAC 2306
 DB 1202 CCGAGAAATATGAATAATGAAGCTGACCAATTAAGAAATACCATTAATGAGCAATATAC 1261

QY 2307 TAGGACTGAGCAATGACAGAGGAAATATCTTTCAAAAAACAGCAGAGTCTTATGC 2366
 DB 1262 TAGGACTGAGCAATGACAGAGGAAATATCTTTCAAAAAACAGCAGAGTCTTATGC 1321

QY 2367 GCTTCCAGTGGATTTACTGAAAAATGAAAAATTTCTGAGAGAGTCAAGGCAACA 2426
 DB 1322 GCTTCCAGTGGATTTACTGAAAAATGAAAAATTTCTGAGAGAGTCAAGGCAACA 1381

QY 2427 TCTGATGGAGCTGACACAGCAGTGAAGTCAAAACCATGACACAGATTAACAAAAAGA 2486
 DB 1382 TCTGATGGAGCTGACACAGCAGTGAAGTCAAAACCATGACACAGATTAACAAAAAGA 1441

QY 2487 AGTCATTAATTAATTTGCGACTGAAAAAT 2516
 DB 1442 AGTCATTAATTAATTTGCGACTGAAAAAT 1471

RESULT 5
 ABA04916
 ID ABA04916 standard; DNA: 1443 BP.

XX ABA04916;
 AC 05-MAR-2002 (first entry)
 XX
 DT
 XX
 DE Human RNA helicase RH116 related DNA sequence #3.
 XX
 KW Human: RH116; RNA helicase; cytosolic; virucide; anti-HIV;
 KW immunosuppressive; immunostimulatory; antithrombotic;
 KW antihypertensive; osteoporosis; antidiabetic; hepatotropic;
 KW antineoplastic; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 11-MAY-2001; 2001WO-FR01441.
 XX
 PR 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 XX (INSP) INST PASTEUR LILLE.
 XX
 PI Bahr G, Cocude C, Capron A;


```
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231966.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0232404.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236370.
XX

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX P-PSDB; AAU23647.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
```


DR P-PSDB: AAU00296.
 XX
 XX New Interferon Induced polypeptides and polynucleotides; useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 XX Claim 9; Page 29-32; 134pp: English.
 XX
 CC The sequence represents Interferon induced nucleic acid, IFN6. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying Interferon-like proteins and Interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 XX Sequence 1258 BP; 399 A; 267 C; 276 G; 316 ?; 0 other:
 S0
 Query Match 19.3%; Score 651; DB 22; Length 1258;
 Best Local Similarity 100.0%; Pred. No. 2,36-303;
 Matches 651: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 CAGAGGTGAAGGACGATTCAGAGGACAGTCCGCGGAAACATTCAGGAGTTG 346
 DB 257 CAGAGGTGAAGGACGATTCAGAGGACAGTCCGCGGAAACATTCAGGAGTTG 316
 QY 347 AACTGCTGAGACCTTGAGGAGGAGTGTGACCTTGAGTGGATTCAGGAGTTG 406
 DB 317 AACTGCTGAGACCTTGAGGAGGAGTGTGACCTTGAGTGGATTCAGGAGTTG 376
 QY 407 TGGAGCCCTCCGAGAACCCGAGCCCTGCGCCCGCGTACATGAATTCAGGAGTTG 466
 DB 377 TGGAGCCCTCCGAGAACCCGAGCCCTGCGCCCGCGTACATGAATTCAGGAGTTG 436
 QY 467 CGGACTTGCCTCCATGCTTTGAGAACGCTCATGATTAATATCTCCATTCGTAACC 526
 DB 437 CGGACTTGCCTCCATGCTTTGAGAACGCTCATGATTAATATCTCCATTCGTAACC 496
 QY 527 TCCCTTACGCGACTGCTGGAGAACGCTTCTAGTAGAACCTTTGGAATAGTGCATGG 586
 DB 497 TCCCTTACGCGACTGCTGGAGAACGCTTCTAGTAGAACCTTTGGAATAGTGCATGG 556
 QY 587 AGGAGAACTGTTGACAAATGGAAGAACCCGATTGCTCTCAGAAATCAATGGA 646
 DB 557 AGGAGAACTGTTGACAAATGGAAGAACCCGATTGCTCTCAGAAATCAATGGA 616
 QY 647 ATGAATCAGGTGTAAGAGAGCTACTAAAGAGATTGTGCAGAAAGAAATTTGTTCTGTG 706
 DB 617 ATGAATCAGGTGTAAGAGAGCTACTAAAGAGATTGTGCAGAAAGAAATTTGTTCTGTG 676
 QY 707 CATTCTGAAATGCTTCTGCTCAACAGAAACAAATGAATTTTCCAAAGTTAAAGGCT 766
 DB 677 CATTCTGAAATGCTTCTGCTCAACAGAAACAAATGAATTTTCCAAAGTTAAAGGCT 736
 QY 767 CTGATTGCTCAAGAAAGCAATGAGAGATTGAGATTATCAACATTTGATTCCTCAAG 826
 DB 737 CTGATTGCTCAAGAAAGCAATGAGAGATTGAGATTATCAACATTTGATTCCTCAAG 796

QY 827 TGAAGAGCAACTTCTTCAACACAGTTCAGCCAAATCTGGAGAGAGGCTGTGTCGCA 886
 DB 797 TGAAGAGCAACTTCTTCAACACAGTTCAGCCAAATCTGGAGAGAGGCTGTGTCGCA 856
 QY 887 TGGAGATTAATCATCATCATCATCTTTGAGATTTCTTCTGTGATTTGAG 937
 DB 857 TGGAGATTAATCATCATCATCATCTTTGAGATTTCTTCTGTGATTTGAG 907

RESULT 11
 ID AAS01151
 ID AAS01151 standard; DNA; 1270 BP.
 XX
 XX AAS01151:
 XX
 XX 12-SEP-2001 (first entry)
 XX
 XX Interferon induced nucleic acid, IFN6.
 XX
 XX Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 XX immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 XX graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 XX human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 XX haematologic disease; chronic neutropenia; myocardial infarction;
 XX neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 XX amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN6; ds.
 XX
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX CDS 1..927
 XX /tag- a
 XX /product- "IFN6"
 XX /note- "Interferon induced polypeptide"
 XX
 XX W0200118208-A2.
 XX
 XX 15-MAR-2001.
 XX
 XX 08-SEP-2000; 2000WO-US24704.
 XX
 XX 08-SEP-1999; 99US-0152921.
 XX 20-OCT-1999; 99US-0160575.
 XX 20-JAN-2000; 2000US-0177104.
 XX 07-SEP-2000; 2000US-0656633.
 XX
 XX (CURA-) CURAGEN CORP.
 XX (BIOJ) BIOGEN INC.
 XX
 XX Peyman JA, Da Silva A, Hochman P, Hsu A;
 XX WPI: 2001-235201/24.
 XX P-PSDB: AAU00296.
 XX
 XX New Interferon Induced polypeptides and polynucleotides; useful for the
 XX diagnosis, prevention and treatment of immunological, cell
 XX proliferative disorders, such as lupus erythematosus, cancer, stroke
 XX and Alzheimer's disease
 XX
 XX Claim 9; Page 40-43; 134pp: English.
 XX
 XX The sequence represents Interferon induced nucleic acid, IFN6. IFN
 XX nucleic acids and polypeptides are useful for treating or preventing a
 XX pathology associated with IFN polypeptide in a human. They are useful for
 XX determining the presence of or predisposition to a disease associated
 XX with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 XX acids, polypeptides and antibodies are useful for diagnosis, prevention
 XX or treatment of variety of immunological and cell proliferative
 XX disorders, such as autoimmune diseases e.g. lupus erythematosus,
 XX immunodeficiency diseases such as acquired immunodeficiency syndrome
 XX (AIDS), graft rejection, viral infections including hepatitis and human
 XX immunodeficiency virus (HIV), immune-mediated glomerulonephritis,

CC haematologic diseases such as aplastic anaemia and chronic neutropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon agonists, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans.

XX Sequence 1270 BP; 401 A; 271 C; 279 G; 319 T; 0 other:

Query Match 19.3%; Score 651; DB 22; Length 1270;

Best Local Similarity 100.0%; Pred. No. 2,3e-303;

Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 CAGAGCTGAAGAGCAGATTTCAGAGCAGTCCGACCTCCGGGAAACATCAGGCACTTG 346
DB 269 CAGAGCTGAAGAGCAGATTTCAGAGCAGTCCGACCTCCGGGAAACATCAGGCACTTG 328
QY 347 AACTGCTGTGAGACCTTGGAGAGGAGTCTGGCACCCTGGTTGGATTCAGGAATTG 406
DB 329 AACTGCTGTGAGACCTTGGAGAGGAGTCTGGCACCCTGGTTGGATTCAGGAATTG 388
QY 407 TGGAGCCCTCCGGAGAACCGGACCTCTGGCCCGCCGGTAATGAATCTGACCTTA 466
DB 389 TGGAGCCCTCCGGAGAACCGGACCTCTGGCCCGCCGGTAATGAATCTGACCTTA 448
QY 467 CGGACTTGGCCCTCCCATCGTTTGGAGAGCCTCATGATATATCTCACTGGTGAAC 526
DB 449 CGGACTTGGCCCTCCCATCGTTTGGAGAGCCTCATGATATATCTCACTGGTGAAC 508
QY 527 TCGTTAGCCCACTGTGTGGACACCTTCTAGTAGAGACCTCTGGATAGTGCATCG 586
DB 509 TCGTTAGCCCACTGTGTGGACACCTTCTAGTAGAGACCTCTGGATAGTGCATCG 568
QY 587 AGGAGACCTGTGACATTTGAAGACAGAAACCGGATCTCTGACAGATACATGGAA 646
DB 569 AGGAGACCTGTGACATTTGAAGACAGAAACCGGATCTCTGACAGATACATGGAA 628
QY 647 ATGAATCAGGTGAAGAGAGCTACTTAAAAAGATTGTGCAGAAASAAAA 136TTCTCTG 706
DB 629 ATGAATCAGGTGAAGAGAGCTACTTAAAAAGATTGTGCAGAAASAAAA 136TTCTCTG 688
QY 707 CATTTCTGAATGTTCTCTCTCAAAACAGAAACATGAACTTGTCAAGA 11AACAGGCT 756
DB 689 CATTTCTGAATGTTCTCTCTCAAAACAGAAACATGAACTTGTCAAGA 11AACAGGCT 748
QY 767 CGATGTCACAGAAACATGAGATTTGAGATTTATCACAGTTGATGCTCCAG 826
DB 749 CGATGTCACAGAAACATGAGATTTGAGATTTATCACAGTTGATGCTCCAG 808
QY 827 TGGAAAGCAGACTTTCTTACACACAGTTCAGCCAAATCTGAGAGAGCA 16CTGGGCA 886
DB 809 TGGAAAGCAGACTTTCTTACACACAGTTCAGCCAAATCTGAGAGAGCA 16CTGGGCA 868
QY 887 TGGAGATTAACATCATCATCTTTTGGAGATTCTCTGATTTTCTG 937
DB 869 TGGAGATTAACATCATCATCTTTTGGAGATTCTCTGATTTTCTG 919

RESULT 12
AADI7207
ID AADI7207 standard; DNA; 6406 BP.
XX AADI7207;
AC AADI7207;
XX 29-NOV-2001 (first entry)
DT 29-NOV-2001 (first entry)
XX Human melanoma differentiation associated-5 gene full length promoter.
DE
XX

KM Human; melanoma differentiation associated gene; Mda-5; Interferon; IFN;
KM RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;
KM neuroblastoma; astrocytoma; glioblastoma; multifocal; cancer; cervical;
KM breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
KM central nervous system; cytosolic; apoptosis; ds.

OS Homo sapiens.

PN WO200164707-A1.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001WO-US06960.

PR 29-FEB-2000; 2000US-0515363.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Fisher PB, Kang D, Gopalakrishnan RV;

DR WPI; 2001-565494/63.

PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
PT gene useful for cancer cell growth suppression, apoptosis and
PT anti-viral activity.

PS Example 3; Page 94-101; 152pp; English.

CC The present invention relates to an isolated nucleic acid encoding a
CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
CC Mda-5 is a novel interferon (IFN) inducible gene with structural
CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
CC is induced during terminal differentiation in human melanoma cells
CC treated with the combination of recombinant fibroblast IFN and the
CC anti-leukaemic compound mezerein (MEZ). Mda-5 is useful for identifying
CC compounds that may induce its expression. Mda-5 is useful for treating
CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
CC multifocal, cervical cancer, breast cancer, colon cancer, prostate
CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
CC The present sequence is human Mda-5 gene full length promoter sequence.

SQ Sequence 6406 BP; 2077 A; 1388 C; 1229 G; 1712 T; 0 other:

Query Match 16.5%; Score 556; DB 22; Length 6406;

Best Local Similarity 100.0%; Pred. No. 1.6e-257;

Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCGGCTGAGAGCCCTGTGACAACTCGTCATTGTGAGCAGACAGCGGTAGAC 60
DB 5851 GCGGCGCGGCTGAGAGCCCTGTGACAACTCGTCATTGTGAGCAGACAGCGGTAGAC 5910
QY 61 CCGTCTTCTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB 5911 CCGTCTTCTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5970
QY 121 CAGCACCATCTGCTGGAGAAACCTCTCTCTCTGAGAAAGAAAGATGTGAAATGG 180
DB 5971 CAGCACCATCTGCTGGAGAAACCTCTCTCTCTGAGAAAGAAAGATGTGAAATGG 6030
QY 181 TATTCACAGAGAGAAATTTCCGCTATCTCATCTGCTCTGAGGCGCAGGATGAAATG 240
DB 6031 TATTCACAGAGAGAAATTTCCGCTATCTCATCTGCTCTGAGGCGCAGGATGAAATG 6090
QY 241 TACATCAGAGTGAAGCTGTGTGAGTACCTGACCTTCTGCTGAGAGAGTGAAGAG 300
DB 6091 TACATCAGAGTGAAGCTGTGTGAGTACCTGACCTTCTGCTGAGAGAGTGAAGAG 6150
QY 301 CAGATTGAGAGAGAGTGGCCACCTCGGGGAAACATGACAGGAGATTGAACCTCTGAGC 360
DB 6151 CAGATTGAGAGAGAGTGGCCACCTCGGGGAAACATGACAGGAGATTGAACCTCTGAGC 6210

QY 361 ACCTGGAGAGAGAGTGGACACCTTGGTTGAGCTCGGGAATTTGGAGTCCCTCCG 420
 Db 6211 ACCTGGAGAGAGAGTGGACACCTTGGTTGAGCTCGGGAATTTGGAGTCCCTCCG 6270
 QY 421 AGAAGCGGCGAGCCCTGGCGCGCGCTACATGAACCTTACGTCAAGTCTTGGCTT 480
 Db 6271 AGAAGCGGCGAGCCCTGGCGCGCGCTACATGAACCTTACGTCAAGTCTTGGCTT 6330
 QY 481 CCATCGTTGAGAGCGCTCATGATGAATATCTCCAACTGCTGAACCTTCCAGCCAT 540
 Db 6331 CCATCGTTGAGAGCGCTCATGATGAATATCTCCAACTGCTGAACCTTCCAGCCAT 6390
 QY 541 CTGTGGACAGCTTC 556
 Db 6391 CTGTGGACAGCTTC 6406

RESULT 13

ABA61000/C
 ID ABA61000 standard; DNA; 595 BP.

AC ABA61000;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #930

KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -

PS Claim 1: SEQ ID NO 9305; 639bp + sequence listing; Engr Sh.

CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 595 BP; 141 A; 162 C; 154 G; 138 T; 0 other;

XX Query Match 15.0%; Score 505; DB 22; Length 595;

XX Best Local Similarity 100.0%; Pred. No. 6; Le-23%;

XX Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 ACAAGACACATCTGCTTGGAGAACCTCTCCCTTCTGTGAGAAAGAAAGATCTCGAA 176
 Db 595 ACAAGACACATCTGCTTGGAGAACCTCTCCCTTCTGTGAGAAAGAAAGATCTCGAA 536
 QY 177 TGGGATTTCCAGAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAGGCCAGGTGAA 236
 Db 535 TGGGATTTCCAGAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAGGCCAGGTGAA 476
 QY 237 AATGTACATCCAG 296
 Db 475 AATGTACATCCAG 416
 QY 297 GGAGCAGATTGAG 356
 Db 415 GGAGCAGATTGAG 356
 QY 357 GAGCAGATTGAG 416
 Db 355 GAGCAGATTGAG 296
 QY 417 CCGAGAACCGGAG 476
 Db 295 CCGAGAACCGGAG 236
 QY 477 CTCTCCATGCTTTGAG 536
 Db 235 CTCTCCATGCTTTGAG 176
 QY 537 CACTCTGGTGAG 596
 Db 175 CACTCTGGTGAG 116
 QY 597 GTTGACAAATTGAAG 621
 Db 115 GTTGACAAATTGAAG 91

RESULT 14

AAK09292/C
 ID AAK09292 standard; DNA; 595 BP.

AC AAK09292;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 9283.

KM Human; brain expressed exon; gene expression analysis; probe;

KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KM epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

XX
PN
XX

W0300157276-A2.

Db 115 GTTGACATTGAGACAGAAACGG 91

Sat May 17 10:42:58 2003

us-09-515-363c-1.rng

Search completed: May 17, 2003, 03:39:19
Job time : 775 secs

• • •